AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

ARTIFICIAL INTELLIGENCE: SOCIETAL AND ETHICAL IMPLICATIONS

HEARING

BEFORE THE

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY
HOUSE OF REPRESENTATIVES

ONE HUNDRED SIXTEENTH CONGRESS

FIRST SESSION

JUNE 26, 2019

Serial No. 116–32

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
36–796PDF WASHINGTON : 2019

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

HON. EDDIE BERNICE JOHNSON, Texas, *Chairwoman*

ZOE LOFGREN, California
DANIEL LIPINSKI, Illinois
SUZANNE BONAMICI, Oregon
AMI BERA, California,
Vice Chair
CONOR LAMB, Pennsylvania
LIZZIE FLETCHER, Texas
HALEY STEVENS, Michigan
KENDRA HORN, Oklahoma
MIKIE SHERRILL, New Jersey
BRAD SHERMAN, California
STEVE COHEN, Tennessee
JERRY MCNERNEY, California
ED PERLMUTTER, Colorado
PAUL TONKO, New York
BILL FOSTER, Illinois
DON BEYER, Virginia
CHARLIE CRIST, Florida
SEAN CASTEN, Illinois
KATIE HILL, California
BEN MCADAMS, Utah
JENNIFER WEXTON, Virginia

FRANK D. LUCAS, Oklahoma,
Ranking Member
MO BROOKS, Alabama
BILL POSEY, Florida
RANDY WEBER, Texas
BRIAN BABIN, Texas
ANDY BIGGS, Arizona
ROGER MARSHALL, Kansas
RALPH NORMAN, South Carolina
MICHAEL CLOUD, Texas
TROY BALDERSON, Ohio
PETE OLSON, Texas
ANTHONY GONZALEZ, Ohio
MICHAEL WALTZ, Florida
JIM BAIRD, Indiana
JAIME HERRERA BEUTLER, Washington
JENNIFFER GONZALEZ-COLÓN, Puerto Rico
VACANCY

C O N T E N T S

June 26, 2019

	Page
Hearing Charter	2

Opening Statements

Statement by Representative Eddie Bernice Johnson, Chairwoman, Committee on Science, Space, and Technology, U.S. House of Representatives	8
Written statement	9
Statement by Representative Jim Baird, Committee on Science, Space, and Technology, U.S. House of Representatives	9
Written statement	11
Written statement by Representative Frank Lucas, Ranking Member, Committee on Science, Space, and Technology, U.S. House of Representatives	11

Witnesses:

Ms. Meredith Whittaker, Co-Founder, AI Now Institute, New York University	
Oral Statement	13
Written Statement	16
Mr. Jack Clark, Policy Director, OpenAI	
Oral Statement	32
Written Statement	34
Mx. Joy Buolamwini, Founder, Algorithmic Justice League	
Oral Statement	45
Written Statement	47
Dr. Georgia Tourassi, Director, Oak Ridge National Lab-Health Data Sciences Institute	
Oral Statement	74
Written Statement	76
Discussion	92

Appendix I: Answers to Post-Hearing Questions

Ms. Meredith Whittaker, Co-Founder, AI Now Institute, New York University	120
Mr. Jack Clark, Policy Director, OpenAI	123
Mx. Joy Buolamwini, Founder, Algorithmic Justice League	128
Dr. Georgia Tourassi, Director, Oak Ridge National Lab-Health Data Sciences Institute	135

Appendix II: Additional Material for the Record

H. Res. 153 submitted by Representative Haley Stevens, Chairwoman, Subcommittee on Research and Technology, Committee on Science, Space, and Technology, U.S. House of Representatives	140

ARTIFICIAL INTELLIGENCE: SOCIETAL AND ETHICAL IMPLICATIONS

WEDNESDAY, JUNE 26, 2019

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

The Committee met, pursuant to notice, at 10 a.m., in room 2318 of the Rayburn House Office Building, Hon. Eddie Bernice Johnson [Chairwoman of the Committee] presiding.

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

HEARING CHARTER

Artificial Intelligence: Societal and Ethical Implications

Wednesday, June 26, 2019
10 a.m. – 12:00 p.m.
2318 Rayburn House Office Building

Purpose:

On Wednesday, June 26, 2019, the Science, Space, and Technology Committee will hold a hearing to discuss the impact of artificial intelligence (AI) on society and the ethical implications in the design and use of this technology. The hearing will examine the extent to which AI is already being deployed across different sectors of our society and economy, how biases, vulnerabilities, and other unintended consequences may manifest in these AI systems, and how Federal agencies, as part of their research programs, standards development efforts, and internal adoption of AI, can help ensure more ethical and responsible design and application of AI.

Witnesses:

- **Ms. Meredith Whittaker**, Co-Founder, AI Now Institute, New York University
- **Mr. Jack Clark,** Policy Director, OpenAI
- **Mx. Joy Buolamwini**, Founder, Algorithmic Justice League
- **Dr. Georgia Tourassi,** Director, Oak Ridge National Lab—Health Data Sciences Institute

Overarching Questions:

- In what applications and to what extent are AI systems already in use today? What are examples of AI use that touch people's lives every day that we don't often hear about?
- What are the different ways that bias can manifest in AI systems? What are the consequences of these biases? What are some of the other risks and concerns related to fairness, transparency, trust and other ethical considerations in the application of AI systems?
- How should we assess and address these risks and concerns in AI systems? How can we integrate ethical considerations at the earliest stages of research and education? What is the role of the Federal government in these efforts?

Background

Ubiquity of AI

All applications of artificial intelligence in use today can be considered "narrow AI," or AI that's designed to do a very specific set of tasks. (In contrast, general AI is a system that possesses generalized human cognitive abilities and, when presented with an unfamiliar and complex problem, can develop solutions drawing from contextual knowledge. We are still very far from achieving artificial general intelligence.) Machine learning is a technique most often used to achieve end-user AI applications, and 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. "bike", "cat", "lamp") can be fed into an algorithm, which then looks for patterns common to all images with a specific label. The algorithm builds a model ("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 is able to 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.

AI systems have been in use for a while in the commercial sector, the most prominent examples being targeted advertising and financial market predictions. More recently, thanks to rapid advances in computing speed and methodology (e.g. deep neural networks), as well as increasingly larger datasets generated and collected across a variety of platforms, AI-powered systems have grown increasingly capable and widespread. In healthcare, AI systems can aid in medical diagnoses[1,2], perform many duties of clinical assistants[3], and help first responders make critical decisions[4]. In transportation, AI algorithms can help predict and mitigate traffic[5], and autonomous vehicles that use a variety of AI technologies are rapidly becoming more advanced[6]. AI technology used in agriculture can improve crop quality and reduce workloads[7], and AI algorithms are increasingly used in scientific research to help sort and analyze massive amounts of data in fields such as weather prediction[8] and genetics research[9]. Businesses large and small

[1] https://www.nytimes.com/2019/02/11/health/artificial-intelligence-medical-diagnosis.html
[2] http://med.stanford.edu/news/all-news/2019/06/researchers-develop-ai-tool-to-help-detect-aneurysms.html
[3] https://www.meridian.edu/how-artificial-intelligence-ai-is-improving-medical-assisting/
[4] https://www.meritalk.com/articles/u-s-canada-ai-partnership-aims-to-help-first-responders/
[5] https://news.usc.edu/148660/usc-engineers-use-artificial-intelligence-to-reduce-traffic-jams/
[6] https://www.ucsusa.org/clean-vehicles/how-self-driving-cars-work
[7] https://searchenterpriseai.techtarget.com/feature/Agricultural-AI-yields-better-crops-through-data-analytics
[8] https://spacenews.com/ai-for-earth-observation-and-numerical-weather-prediction/
[9] https://www.deepgenomics.com/

are increasingly adopting AI technology to improve performance quality and workflow efficiency—AI analysis has even been used to improve beer brewing[10] and clean cat litter[11].

AI Associated Risks

AI-powered systems have the potential to drastically improve our lives, but also the potential to do significant harm if they are not vetted for bias and fairness. (This hearing is primarily focused on civilian and commercial uses of AI with a presumption of no intent to harm. There are many scenarios in which AI can be intentionally misused or abused.) There are many different causes and manifestations of bias, the most straightforward of which is bias in training data. An AI algorithm's performance depends heavily on the quality of its training data. In the image recognition example above, if the tagged training dataset included mostly cats but only a few dogs, the algorithm will be able to identify cats much better than dogs. In more practical examples, 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. Training data bias can have significant social implications as well. Facial recognition systems trained on mostly light-skinned faces have performed much worse in identifying faces with darker skin[12]. When such systems are used in law enforcement to, for example, identify criminal suspects from video footage, it can lead to a higher number of false arrests for people with darker skin.

One solution to this problem can be to "de-bias" the data by making sure the data is representative of real life. However, such an approach can quickly exacerbate societal biases, because real life data reflect existing social norms and structures of power. Targeted job advertising services have shown to men advertisements related to higher paying jobs than what is shown to women[13]. When a user searches black-identifying names in Google, they are more likely to see arrest-related ads than when searching white-identifying names[14]. Even when AI systems are specifically designed to mitigate human bias, hidden biases can arise. A well-known example is Amazon's attempt to build a resume screening AI algorithm to identify promising job candidates[15]. Part of the goal was to eliminate personal bias from human hiring managers, who might rate applicants higher if the manager relates to them more or if the candidate fits the manager's subjective standards of qualification. The algorithm was trained using resumes submitted to the company over a 10-year period. However, because the tech sector has been severely male dominated over the past decade, the algorithm quickly learned that male candidates were more preferable and demoted any resume that mentions the word "women" or

[10] https://www.forbes.com/sites/bernardmarr/2019/02/01/how-artificial-intelligence-is-used-to-make-beer/#904f8e370cf4
[11] https://www.techradar.com/news/this-ai-litter-tray-analyzes-your-cats-health-and-uses-nasa-tech-to-clean-itself
[12] https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html
[13] https://www.theguardian.com/technology/2015/jul/08/women-less-likely-ads-high-paid-jobs-google-study
[14] https://www.technologyreview.com/s/510646/racism-is-poisoning-online-ad-delivery-says-harvard-professor/
[15] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

indicates that the applicant was female (e.g. "captain of softball team" vs "quarterback"), even though the AI was not explicitly programmed to consider gender. Amazon eventually cancelled this project, but if the algorithm had been implemented in real life without being vetted for bias, it would have exacerbated the already significant gender inequality in the tech sector. If a dataset set is carefully curated and vetted for bias and fairness, it could solve some of the issues associated with biases manifesting in AI systems.

There are additional sources of bias that can be introduced in the design phase before an algorithm is ever trained on data. For instance, AI algorithms can be designed to optimize for a small set of parameters without considering the bigger context of the problem. An extreme example is, if an AI is tasked with developing a method to suppress a widespread disease, it might propose to eradicate an entire country's population. In this case, the AI optimized only for disease control without regard to the broader context of the goal, which is to save human lives. Bias can also arise when a measured characteristic is used as a poor proxy for another characteristic. For example, risk assessment algorithms are increasingly used in courts to determine bail or even jail time by evaluating factors such as gender, age, and prior convictions. However, the extent to which each of these characteristics contribute to a person's likelihood to commit a crime is still an active area of research, and therefore the risk assessments already in use are not clearly based on sound science.

These above instances are examples of poor alignment between the task assigned to the AI and the actual human goal. To better align the AI tasks and human goals involves not only technology expertise, but an understanding of the relevant social science and ethical considerations. Bias is not a technical bug, but a social problem. Because humans program AI, the programmers' biases can naturally carry through into the AI system, and it requires an interdisciplinary approach to mitigate these biases.

When AI systems are shown to produce biased results, the systems may be re-trained or re-designed to produce more equitable outcomes. However, biases may remain hidden in the AI "black box." In addition, many users of AI-driven products may lack the awareness and expertise to test for bias or fairness before implementing AI systems– i.e. they may have undue trust in the system. Finally, when biased AI systems are put into applications such as in criminal justice, schools, or financial sectors, the technology can discriminate against many more people and much faster than any one biased individual can, exacerbating existing inequities and perhaps creating new ones. All of these risks are greater when humans are out of the decision-making loop and there is no opportunity for the affected individuals to appeal the AI's "decision" – i.e. there is a lack of transparency[16]. Beyond any one application or algorithm, experts have also raised broader questions of who benefits from AI more so than others, if the widespread deployment of AI could further exacerbate existing inequities due to job loss and disparate access to the benefits of AI, and whether AI tools should be used at all in certain contexts.

[16] https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy

Ethical Design and Deployment of AI

Well before AI systems are deployed in our society, there are many ways in which ethical considerations can be integrated in the research and design processes, as well as in the education and training of the scientists and engineers who will ultimately design these systems.

Researchers deciding on what research questions to pursue and what kinds of systems to design can engage in the exercise of imagining every application – good and bad- to which the research or algorithm may be relevant and every way in which biases may manifest. There is no way to predict all such possible outcomes, but the very exercise of considering possibilities encourages researchers to put their research in a societal context and refine their path for the best possible outcomes.

Computer and data scientists can partner with ethicists, social scientists, legal scholars, and others in the humanities and social sciences to bring to bear their scholarly expertise and perspective in shaping research and designing systems. Diversity in personal experience and perspective is critical to minimizing bias. The representation of women and minorities in AI research and the tech sector more broadly is already very poor, as reported widely in recent years. This lack of diversity in those designing systems manifests in the technology in unintended ways, such as in the Amazon example provided previously.

Real-world environments for AI applications are almost always different from lab settings. Users of AI systems, especially public sector users such as schools, police, courts, and others can rigorously test their systems at the point of use and actively engage with the public in that process to uncover hidden or overlooked biases.

Achieving the responsible design and deployment of AI also requires integrating ethics into technology education at every stage of the AI education pipeline, from K-12 all the way up to current AI developers. It requires viewing AI as an interdisciplinary field rather than a purely technical field. The National Science Foundation (NSF), which funds university research across all non-biomedical disciplines (including social sciences) and also funds numerous STEM education programs, has a critical role in both of these efforts.

Standards around training datasets, performance measures, and best practices for assessing the impact of AI systems could help current AI developers and users design and use AI more responsibly. The National Institute for Standards and Technology (NIST) has begun broad stakeholder engagement in thinking about what standards and frameworks around AI could look like, as part of complying with the Executive Order on Artificial Intelligence[17]. This includes

[17] https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/

holding a workshop with relevant stakeholder[18] and issuing a Request For Information (RFI) regarding AI standards[19].

Many universities and think tanks are already considering the ethical issues surrounding AI R&D. For example, Stanford recently established its Institute for Human Centered AI (Stanford HAI)[20], which aims to take a multidisciplinary approach to AI research by bringing together faculty and researchers from across the university campus. The Harvard Berkman Klein Center and the MIT Media Lab have partnered to create the Assembly program[21], which brings together technologists, business managers, and policymakers to tackle emerging problems related to the ethics and governance of AI. The private sector is also attempting to tackle issues related to AI bias and ethics. Companies such as Microsoft, Google, and Intel have all published their own versions of AI ethics principles[22, 23, 24]. However, these principles are generally abstract and lack concrete governance structures and accountability measures.

Finally, there are also international conversations taking place surrounding the ethics of AI. The Organisation for Economic Cooperation and Development (OECD) recently adopted a set of AI principles for guiding governments in responsible stewardship of trustworthy AI[25]. Many individual countries have also established their own AI strategies that incorporate ethics to various extents. However, similar to private companies' attempts to address AI ethics, many of these plans and principles are high level and abstract, and more concrete steps are still in development.

---

[18] https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/
[19] https://www.federalregister.gov/documents/2019/05/01/2019-08818/artificial-intelligence-standards
[20] https://hai.stanford.edu/
[21] https://bkmla.org/
[22] https://www.microsoft.com/en-us/ai/our-approach-to-ai
[23] https://ai.google/responsibilities/responsible-ai-practices/
[24] https://newsroom.intel.com/articles/intels-recommendations-u-s-national-strategy-artificial-intelligence/#gs.jzrn1u
[25] https://www.oecd.org/going-digital/ai/principles/

Chairwoman JOHNSON. The hearing will come to order. Without objection, the Chair is authorized to declare recess at any time.

Good morning, and welcome to our distinguished panel of witnesses. We are here today to learn about the societal impacts and ethical implications of a technology that is rapidly changing our lives, namely, artificial intelligence. From friendly robot companions to hostile terminators, artificial intelligence (AI) has appeared in films and sparked our imagination for many decades.

Today, it is no longer a futuristic idea, at least not artificial intelligence designed for a specific task. Recent advances in computing power and increases in data production and collection have enabled artificial-intelligence-driven technology to be used in a growing number of sectors and applications, including in ways we may not realize. It is routinely used to personalize advertisements when we browse the internet. It is also being used to determine who gets hired for a job or what kinds of student essays deserve a higher score.

The artificial intelligence systems can be a powerful tool for good, but they also carry risk. The systems have been shown to exhibit gender discrimination when displaying job ads, racial discrimination in predictive policing, and socioeconomic discrimination when selecting zip codes to offer commercial products or services.

The systems do not have an agenda, but the humans behind the algorithms can unwittingly introduce their personal biases and perspectives into the design and use of artificial intelligence. The algorithms are then trained with data that is biased in ways both known and unknown. In addition to resulting in discriminatory decisionmaking, biases in design and training of algorithms can also cause artificial intelligence to fail in other ways, for example, performing worse than clinicians in medical diagnostics. We know that these risks exist. What we do not fully understand is how to mitigate them.

We are also struggling with how to protect society against intended misuse and abuse. There has been a proliferation of general artificial intelligence ethics principles by companies and nations alike. The United States recently endorsed an international set of principles for the responsible development. However, the hard work is in the translation of these principles into concrete, effective action. Ethics must be integrated into the earliest stages of the artificial intelligence research and education, and continue to be prioritized at every stage of design and deployment.

Federal agencies have been investing in this technology for years. The White House recently issued an executive order on Maintaining American Leadership in artificial intelligence and updated the 2016 National Artificial Intelligence R&D Strategic Plan. These are important steps. However, I also have concerns. First, to actually achieve leadership, we need to be willing to invest. Second, while few individual agencies are making ethics a priority, the Administration's executive order and strategic plan fall short in that regard. When mentioning it at all, they approach ethics as an add-on rather than an integral component of all artificial intelligence R&D (research and development).

From improving healthcare, transportation, and education, to helping to solve poverty and improving climate resilience, artificial

intelligence has vast potential to advance the public good. However, this is a technology that will transcend national boundaries, and if the U.S. does not address the ethics seriously and thoughtfully, we will lose the opportunity to become a leader in setting the international norms and standards in the coming decades. Leadership is not just about advancing the technology; it is about advancing it responsibly.

I look forward to hearing the insights and recommendation from today's expert panel on how the United States can lead in the ethical development of artificial intelligence.

[The prepared statement of Chairwoman Johnson follows:]

Good morning, and welcome to our distinguished panel of witnesses.

We are here today to learn about the societal impacts and ethical implications of a technology that is rapidly changing our lives, namely, Artificial intelligence.

From friendly robot companions to hostile terminators, AI has appeared in films and sparked our imagination for many decades. Today, AI is no longer a futuristic idea, at least not AI designed for specific tasks. Recent advances in computing power and increases in data production and collection have enabled AI-driven technology to be used in a growing number of sectors and applications, including in ways we may not realize. AI is routinely used to personalize advertisements when we browse the internet. It is also being used to determine who gets hired for a job or what kinds of student essays deserve a higher score.

AI systems can be a powerful tool for good, but they also carry risks. AI systems have been shown to exhibit gender discrimination when displaying job ads, racial discrimination in predictive policing, and socioeconomic discrimination when selecting which zip codes to offer commercial products or services.

The AI systems do not have an agenda, but the humans behind the algorithms can unwittingly introduce their personal biases and perspectives into the design and use of AI. The algorithms are then trained with data that is biased in ways both known and unknown. In addition to resulting in discriminatory decision-making, biases in the design and training of algorithms can also cause AI to fail in other ways, for example performing worse than clinicians in medical diagnostics.

We know that these risks exist. What we do not fully understand is how to mitigate them. We are also struggling with how to protect society against intended misuse and abuse of AI. There has been a proliferation of general AI ethics principles by companies and nations alike. The United States recently endorsed an international set of principles for the responsible development of AI. However, the hard work is in the translation of these principles into concrete, effective action. Ethics must be integrated at the earliest stages of AI research and education, and continue to be prioritized at every stage of design and deployment.

Federal agencies have been investing in AI technology for years. The White House recently issued an executive order on Maintaining American Leadership in AI and updated the 2016 National Artificial Intelligence R&D Strategic Plan. These are important steps. However, I also have concerns. First, to actually achieve leadership, we need to be willing to invest. Second, while a few individual agencies are making ethics a priority, the Administration's executive order and strategic plan fall short in that regard. When mentioning it at all, they approach ethics as an add-on rather than an integral component of all AI R&D.

From improving healthcare, transportation, and education, to helping to solve poverty and improving climate resilience, AI has vast potential to advance the public good. However, this is a technology that will transcend national boundaries, and if the U.S. does not address AI ethics seriously and thoughtfully, we will lose the opportunity to become a leader in setting the international norms and standards for AI in the coming decades. Leadership is not just about advancing the technology, it's about advancing it responsibly.

I look forward to hearing the insights and recommendations from today's expert panel on how the United States can lead in the ethical development of AI.

Chairwoman JOHNSON. I now recognize Mr. Baird for his opening statement.

Mr. BAIRD. Thank you, Chairwoman Johnson, for holding this hearing today on the societal and ethical implications of artificial intelligence, AI.

In the first half of the 20th century, the concept of artificial intelligence was the stuff of science fiction. Today, it's a reality. Since the term AI was first coined in the 1950s, we have made huge advances in the field of artificial narrow intelligence. Narrow AI systems can perform a single task like providing directions through Siri or giving you weather forecasts. This technology now touches every part of our lives and every sector of the economy.

Driving the growth of AI is the availability of big data. Private companies and government have collected large datasets, which, combined with advanced computing power, provide the raw material for dramatically improved machine-learning approaches and algorithms. How this data is collected, used, stored, secured is at the heart of the ethical and policy debate over the use of AI.

AI has already delivered significant benefits for U.S. economic prosperity and national security, but it has also demonstrated a number of vulnerabilities, including the potential to reinforce existing social issues and economic imbalances.

As we continue to lead the world in advanced computing research, a thorough examination of potential bias, ethics, and reliability challenges of AI is critical to maintaining our leadership in technology. The United States must remain the leader in AI, or we risk letting other countries who don't share our values drive the standards for this technology. To remain the leader in AI, I also believe Americans must understand and trust how AI technologies will use their data.

The Trump Administration announced earlier this year an executive order on "Maintaining American Leadership in Artificial Intelligence." Last week, the Administration's Select Committee on AI released a report that identifies its priorities for federally funded AI research. I'm glad that the Administration is making AI research a priority. This is an effort that is going to require cooperation between industry, academia, and Federal agencies. In government, these efforts will be led by agencies under the jurisdiction of this Committee, including NIST (National Institute of Standards and Technology), NSF (National Science Foundation), and DOE (Department of Energy).

We will learn more about one of those research efforts from one of our witnesses today, Dr. Georgia Tourassi, the Founding Director of the Health Data Sciences Institute at Oak Ridge National Laboratory. Dr. Tourassi's research focuses on deploying AI to provide diagnoses and treatment for cancer. Her project is a good example of how cross-agency collaboration and government data can responsibly drive innovation for public good. I look forward to hearing more about her research.

Over the next few months, this Committee will be working toward bipartisan legislation to support a national strategy on artificial intelligence. The challenges we must address are how industry, academia, and the government can best work together on AI challenges, including ethical and societal questions, and what role the Federal Government should play in supporting industry as it drives innovation.

I want to thank our accomplished panel of witnesses and their testimony today, and I look forward to hearing what role Congress should play in facilitating this conversation.

[The prepared statement of Mr. Baird follows:]

Chairwoman Johnson, thank you for holding today's hearing on the societal and ethical implications of artificial intelligence (AI).

In the first half of the 20th century, the concept of artificial intelligence was the stuff of science fiction. Today it is reality.

Since the term AI was first coined in the 1950s, we have made huge advances in the field of artificial *narrow* intelligence.

Narrow AI systems can perform a single task like providing directions through Siri or giving you weather forecasts. This technology now touches every part of our lives and every sector of the economy.

Driving the growth of AI is the availability of big data. Private companies and government have collected large data sets, which, combined with advanced computing power, provide the raw material for dramatically improved machine learning approaches and algorithms.

How this data is collected, used, stored, and secured is at the heart of the ethical and policy debate over the use of AI.

AI has already delivered significant benefits for U.S. economic prosperity and national security.

But it has also demonstrated a number of vulnerabilities, including the potential to reinforce existing social issues and economic imbalances.

As we continue to lead the world in advanced computing research, a thorough examination of potential bias, ethics, and reliability challenges of AI is critical to maintaining our leadership in this technology.

The United States must remain the leader in AI, or we risk letting other countries who don't share our values drive the standards for this technology.

To remain the leader AI, I believe Americans must also understand and trust how AI technologies will use their data.

The Trump Administration announced earlier this year an Executive Order on ″Maintaining American Leadership in Artificial Intelligence.″

Last week the Administration's Select Committee on AI released a report that identifies its priorities for federally funded AI research.

I am glad that the Administration is making AI research a priority.

This is an effort that is going to require cooperation between industry, academia and federal agencies.

In government, these efforts will be led by agencies under the jurisdiction of this Committee, including NIST, NSF and DOE.

We will learn more about one of those research efforts from one of our witnesses today, Dr. Georgia Tourassi, the founding Director of the Health Data Sciences Institute (HDSI) at Oak Ridge National Laboratory. Dr. Tourassi's research focuses on deploying AI to provide diagnoses and treatment of cancer.

Her project is a good example of how cross-agency collaboration and government data can responsibly drive innovation for public good. I look forward to hearing more about her research.

Over the next few months, this Committee will be working towards bipartisan legislation to support a national strategy on Artificial Intelligence.

The challenges we must address are how industry, academia, and the government can best work together on AI challenges, including ethical and societal questions, and what role the federal government should play in supporting industry as it drives innovation.

I want to thank our accomplished panel of witnesses for their testimony today and I look forward to hearing what role Congress should play in facilitating this conversation.

Chairwoman JOHNSON. Thank you very much.

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 Mr. Lucas follows:]

Today, we will explore the various applications and societal implications of Artificial Intelligence (AI), a complex field of study where researchers train computers to learn directly from information without being explicitly programmed - like humans do.

Last Congress, this Committee held two hearings on this topic - examining the concept of Artificial General Intelligence (AGI) and discussing potential applications for AI development through scientific machine learning, as well as the cutting-edge basic research it can enable.

This morning we will review the types of AI technologies being implemented all across the country and consider the most appropriate way to develop fair and responsible guidelines for their use.

From filtering your inbox for spam to protecting your credit card from fraudulent activity, AI technologies are already a part of our everyday lives. AI is integrated into every major U.S. economic sector, including transportation, health care, agriculture, finance, national defense, and space exploration.

This influence will only expand. In 2016, the global AI market was valued at over $4 billion and is expected to grow to $169 billion by 2025. Additionally, there are estimates that AI could add $15.7 trillion to global GDP by 2030.

Earlier this year, the Trump Administration announced a plan for ″Maintaining American Leadership in Artificial Intelligence.″

Last week, the Administration's Select Committee on Artificial Intelligence released a report that identifies its priorities for federally funded AI research. These include developing effective methods for human-AI collaboration, understanding and addressing the ethical, legal, and societal implications of AI, ensuring the safety and security of AI systems, and evaluating AI technologies through standards and benchmarks.

Incorporating these priorities while driving innovation in AI will require cooperation between industry, academia, and the Federal government. These efforts will be led by agencies under the jurisdiction of this Committee: the National Institute on Standards and Technology (NIST), the National Science Foundation (NSF), and the Department of Energy (DOE).

The AI Initiative specifically directs NIST to develop a federal plan for the development of technical standards in support of reliable, robust, and trustworthy AI technologies. NIST plans to support the development of these standards by building research infrastructure for AI data and standards development and expanding ongoing research and measurement science efforts to promote adoption of AI in the marketplace.

At the NSF, federal investments in AI span fundamental research in machine learning, along with the security, robustness, and explainability of AI systems. NSF also plays an essential role in supporting diverse STEM education, which will provide a foundation for the next generation AI workforce. NSF also partners with U.S. industry coalitions to emphasize fairness in AI, including a program on AI and Society which is jointly supported by the Partnership on AI to Benefit People and Society (PAI).

Finally, with its world-leading user facilities and expertise in big data science, advanced algorithms, and high-performance computing, DOE is uniquely equipped to fund robust fundamental research in AI.

Dr. Georgia Tourassi, the founding Director of the Health Data Sciences Institute (HDSI), joins us today from Oak Ridge National Laboratory (ORNL) - a DOE Office of Science Laboratory. Dr. Tourassi's research focuses on deploying AI to provide diagnoses and treatment for cancer.

The future of scientific discovery includes the incorporation of advanced data analysis techniques like AI. With the next generation of supercomputers, including the exascale computing systems that DOE is expected to field by 2021, American researchers will be able to explore even bigger challenges using AI. They will have greater power, and even more responsibility.

Technology experts and policymakers alike have argued that without a broad national strategy for advancing AI, the U.S. will lose its narrow global advantage. With increasing international competition in AI and the immense potential for these technologies to drive future technological development, it's clear the time is right for the federal government to lead these conversations about AI standards and guidelines.

I look forward to working with Chairwoman Johnson and the members of the Committee over the next few months to develop legislation that supports this national effort.

I want to thank our accomplished panel of witnesses for their testimony today and I look forward to receiving their input.

Chairwoman JOHNSON. At this time, I will introduce our witnesses. Our first witness is Ms. Meredith Whittaker. Ms. Whittaker is a distinguished research scientist at New York University and Co-Founder and Co-Director of the AI Now Institute, which is dedicated to researching the social implications of artificial intelligence and related technologies. She has over a decade of experience working in the industry, leading product and engineering teams.

Our next witness is Mr. Jack Clark. Mr. Clark is the Policy Director of OpenAI where his work focuses on AI policy and strategy. He's also a Research Fellow at the Center for Security and Emerging Technology at Georgetown University and a member of the Center of the New American Security task force at AI National Security. Mr. Clark also helps run the AI Index, an initiative from the Stanford One Hundred Year Study on AI to track AI progress.

After Mr. Clark is Mx. Joy Buolamwini, who is Founder of the Algorithmic Justice League and serves on the Global Tech Panel convened by the Vice President of the European Union to advise leaders and technology executives on ways to reduce the potential harms of AI. She is also a graduate researcher at MIT where her research focuses on algorithmic bias and computer version systems.

Our last witness, Dr. Georgia Tourassi. Dr. Tourassi is the Founding Director of the Health and Data Sciences Institute and Group Leader of Biomedical Sciences, Engineering, and Computing at the Oak Ridge National Laboratory. Her research focuses on artificial intelligence for biomedical applications and data-driven biomedical discovery. Dr. Tourassi also serves on the FDA (Food and Drug Administration) Advisory Committee and Review Panel on Computer-aided Diagnosis Devices.

Our witnesses should know that you will have 5 minutes for your spoken testimony. Your written testimony will be included in the record for the hearing. When you all have completed your spoken testimony, we will begin with a round of questions. Each Member will have 5 minutes to question the panel.

We now will start with Ms. Whittaker.

TESTIMONY OF MEREDITH WHITTAKER, CO-FOUNDER, AI NOW INSTITUTE, NEW YORK UNIVERSITY

Ms. WHITTAKER. Chairwoman Johnson, Ranking Member Baird, and Members of the Committee, thank you for inviting me to speak today. My name is Meredith Whittaker, and I'm the Co-Founder of the AI Now Institute at New York University. We're the first university research institute dedicated to studying the social implications of artificial intelligence and algorithmic technologies.

The role of AI in our core social institutions is expanding. AI is shaping access to resources and opportunity both in government and in the private sector with profound implications for hundreds of millions of Americans. These systems are being used to judge who should be released on bail; to automate disease diagnosis; to hire, monitor, and manage workers; and to persistently track and surveil using facial recognition. These are a few examples among hundreds. In short, AI is quietly gaining power over our lives and institutions, and at the same time AI systems are slipping farther away from core democratic protections like due process and a right refusal.

In light of this, it is urgent that Congress act to ensure AI is accountable, fair, and just because this is not what is happening right now. We at AI Now, along with many other researchers, have documented the ways in which AI systems encode bias, produce harm, and differ dramatically from many of the marketing claims made by AI companies.

Voice-recognition hears masculine sounding voices better than feminine voices. Facial recognition fails to see black faces and transgendered faces. Automated hiring systems discriminate against women candidates. Medical diagnostic systems don't work for dark-skinned patients. And the list goes on, revealing a persistent pattern of gender and race-based discrimination, among other forms of identity.

But even when these systems do work as intended, they can still cause harm. The application of 100 percent accurate AI to monitor, track, and control vulnerable populations raises fundamental issues of power, surveillance, and basic freedoms in our democratic society. This reminds us that questions of justice will not be solved simply by adjusting a technical system.

Now, when regulators, researchers, and the public seek to understand and remedy potential harms, they're faced with structural barriers. This is because the AI industry is profoundly concentrated, controlled by just a handful of private tech companies who rely on corporate secrecy laws that make independent testing and auditing nearly impossible.

This also means that much of what we do know about AI is written by the marketing departments of these same companies. They highlight hypothetical benevolent uses and remain silent about the application of AI to fossil fuel extraction, weapons development, mass surveillance, and the problems of bias and error. Information about the darker side of AI comes largely thanks to researchers, investigative journalists, and whistleblowers.

These companies are also notoriously non-diverse. AI Now conducted a year-long study of diversity in the AI industry, and the results are bleak. To give an example of how bad it is, in 2018 the share of women in computer science professions dropped below 1960 levels. And this means that women, people of color, gender minorities, and others are excluded from shaping how AI systems function, and this contributes to bias.

Now, while the costs of such bias are borne by historically marginalized people, the benefits of such systems, from profits to efficiency, accrue primarily to those already in positions of power. This points to problems that go well beyond the technical. We must ask who benefits from AI, who is harmed, and who gets to decide? This is a fundamental question of democracy.

Now, in the face of mounting criticism, tech companies are adopting ethical principles. These are a positive start, but they don't substitute for meaningful public accountability. Indeed, we've seen a lot of P.R., but we have no examples were such ethical promises are backed by public enforcement.

Congress has a window to act, and the time is now. Powerful AI systems are reshaping our social institution in way—institutions in ways we're unable to measure and contest. These systems are developed by a handful of private companies whose market interests don't always align with the public good and who shield themselves from accountability behind claims of corporate secrecy. When we are able to examine these systems, too often we find that they are biased in ways that replicate historical patterns of discrimination. It is imperative that lawmakers regulate to ensure that these sys-

tems are accountable, accurate, contestable, and that those most at risk of harm have a say in how and whether they are used.

So in pursuit of this goal, AI Now recommends that lawmakers, first, require algorithmic impact assessments in both public and private sectors before AI systems are acquired and used; second, require technology companies to waive trade secrecy and other legal claims that hinder oversight and accountability mechanisms; third, require public disclosure of AI systems involved in any decisions about consumers; and fourth, enhance whistleblower protections and protections for conscientious objectors within technology companies.

Thank you, and I welcome your questions.

[The prepared statement of Ms. Whittaker follows:]

United States House of Representatives
Committee on Science, Space, and Technology

"Artificial Intelligence: Societal and Ethical Implications"
June 26, 2019

Written Testimony of
Meredith Whittaker
Co-founder and Co-director, AI Now Institute, New York University

Chairwoman Johnson, Ranking Member Lucas, and members of the Committee, thank you for inviting me to speak today. My name is Meredith Whittaker and I am the Co-founder of the AI Now Institute at New York University. AI Now is the first university research institute dedicated to studying the social implications of artificial intelligence and algorithmic technologies ("AI").[1] Our work examines the rapid proliferation of AI systems through core social domains such as criminal justice, health care, employment, and education. In particular we focus on concerns in the areas of bias and inclusion, safety and critical infrastructure, rights and liberties, and labor. As we identify problems in each of these spaces, we work to address them through robust research investigations, community engagement, and key policy interventions.

When Kate Crawford and I founded AI Now in 2016, we were just beginning to see both the extreme promises of AI as well as the extreme risks. Our annual reports have chronicled these risks, issues, and concerns, and I can say with confidence that they have only increased over time.[2] From education, to healthcare, to law enforcement, to hiring and worker management,

---

[1] *About*, AI Now Institute, https://ainowinstitute.org/about.html/.

[2] Kate Crawford et al., The AI Now Report 2016 (Sept. 2016), https://ainowinstitute.org/AI_Now_2016_Report.html; AI Now Inst., The AI Now Report 2017 (2017), https://ainowinstitute.org/AI_Now_2017_Report.html; Nicolas Suzor, Tess Van Geelen & Sarah Myers West, *Evaluating the Legitimacy of Platform Governance: A Review of Research and a Shared Research Agenda*, International Communication Gazette (Feb. 2018), 80(4), pp. 385-400, https://eprints.qut.edu.au/112749/3/112749a.pdf; AI Now Inst., Litigating Algorithms: Challenging Government Use of Algorithmic Decision Systems (Sept. 2018), https://ainowinstitute.org/litigatingalgorithms.html; Kate Crawford & Vladan Joler, *Anatomy of an AI System: The Amazon Echo As an Anatomical Map of Human Labor, Data and Planetary Resources*, AI Now Inst. & Share Lab (Sept. 7, 2018), https://anatomyof.ai; AI Now Inst., Algorithmic Accountability Policy Toolkit (Oct. 2018), https://ainowinstitute.org/aap-toolkit.pdf; https://ainowinstitute.org/litigatingalgorithms.pdf; Meredith Whittaker et al., The AI Now Report 2018 (Dec. 2018), https://ainowinstitute.org/AI_Now_2018_Report.pdf; Rashida Richardson, Jason M. Schultz & Kate Crawford, *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 192 (April 2019), available at https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson-Schultz-Crawford.pdf; Sarah Myers West, Meredith Whittaker & Kate Crawford, Discriminating Systems: Gender, Race and Power in AI (April 2019), https://ainowinstitute.org/discriminatingsystems.html.

and well beyond, the role of AI in our core social institutions is expanding, both in government and the private sector. AI is shaping access to resources and opportunity, with profound implications for hundreds of millions of Americans. These systems are being used to judge who should be released on bail;[3] to automate disease diagnosis in patients;[4] to amplify mass surveillance efforts;[5] and to hire, monitor and manage workers.[6] [7] These are only a handful of examples that provide a view of the power that AI is quietly gaining over our lives and institutions.

At the same time that AI systems are proliferating and concentrating the power to impact our lives, they are slipping further away from core democratic protections such as due process and other forms of accountability. Embedded within technological and legal "black boxes," AI systems raise many questions and provide few true answers.

In light of this, it is urgent that we address this accountability gap and ensure these technologies advance the values of fairness and justice that this institution and many others are dedicated to upholding.

The Emerging Problem Space

Thanks to researchers and investigative journalists, we have significant and alarming evidence of the way in which AI systems encode bias, produce harm, accelerate environmental degradation and the climate crisis,[8] [9] avoid established accountability processes, and differ dramatically from many of the marketing claims made by AI companies.

Take for example the simple act of listening. Many AI systems, such as Amazon's Echo device are constantly recording every sound within range of their microphones. This is, in theory, so they can respond to user voice commands. But they are equally capable of other, unauthorized

---

[3] Sam Levin, *Imprisoned by Algorithms: The Dark Side of California Ending Cash Bail*, The Guardian (Sept. 7, 2018), https://www.theguardian.com/us-news/2018/sep/07/imprisoned-by-algorithms-the-dark-side-of-california-ending-cash-bail/.

[4] *Seeing Potential*, Google Stories (2018), https://about.google/intl/en/stories/seeingpotential/.

[5] Russell Brandom, *Facial Recognition Is Coming to US Airports, Fast-Tracked by Trump*, The Verge (Apr. 18, 2017), https://www.theverge.com/2017/4/18/15332742/us-border-biometric-exit-facial-recognition-scanning-homeland-security/.

[6] Terena Bell, *This Bot Judges How Much You Smile During Your Job Interview*, Fast Company (Jan. 15, 2019), https://www.fastcompany.com/90284772/this-bot-judges-how-much-you-smile-during-your-job-interview/.

[7] Kevin Roose, *A Machine May Not Take Your Job, but One Could Become Your Boss*, N.Y. Times (June 23, 2019), https://www.nytimes.com/2019/06/23/technology/artificial-intelligence-ai-workplace.html/.

[8] anatomyof.ai

[9] Emma Strubell, Ananya Ganesh, Andrew McCallum, *Energy and Policy Considerations for Deep Learning in NLP*, In the 57th Annual Meeting of the Association for Computational Linguistics (Jun. 5 2019), https://arxiv.org/abs/1906.02243.

recordings.[10] Even when welcomed into our homes, voice recognition systems are far from objective in what they "hear" and whose voices count. For example, some systems have been shown to recognize more masculine sounding voices better than feminine voices;[11] such bias also exists in facial recognition systems that fail to recognize black and transgendered faces;[12] [13] automated hiring systems that discriminate against women candidates; medical diagnostic systems don't work for dark skinned patients;[14] sentencing algorithms that discriminate against black defendants; and the list goes on.

But even when these systems don't explicitly encode bias, they can still cause harm. For example, the ACLU tested the use of Amazon's Facial Recognition system, finding that 28 members of Congress were falsely matched with mugshots of those previously arrested for a crime.[15] The problems raised by the application of facial recognition and other AI systems won't be solved by ensuring the technology is 100% accurate. The application of AI to monitor, track, and control vulnerable populations raises fundamental issues, reminding us that questions of justice will not be solved simply by adjusting a technical system.

Even in the face of mounting evidence, the rapid integration of AI into sensitive social domains continues. Government agencies are increasingly using AI and algorithmic systems to assess beneficiaries of social services and manage benefit allocation. In many cases, the outcome of these experiments has been harmful and even deadly to the people such programs are meant to serve.[16] For example, several states have turned to automation for Medicaid benefit allocation. In many cases, these systems have failed due to flaws in the system itself, resulting in serious harm and multi-million dollar lawsuits.[17] In Arkansas, such a system was used to

---

[10] Niraj Chokshi, *Is Alexa Listening? Amazon Echo Sent Out Recording of Couple's Conversation*, N.Y. Times (May 25, 2018), https://www.nytimes.com/2018/05/25/business/amazon-alexa-conversation-shared-echo.html/.

[11] Rachael Tatman, *Google's Speech Recognition Has a Gender Bias*, Making Noise and Hearing Things (July 12, 2016), https://makingnoiseandhearingthings.com/2016/07/12/googles-speech-recognition-has-a-gender-bias/

[12] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, Gender Shades (2018), gendershades.org.

[13] Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots*, ACLU (July 26, 2018), https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28/.

[14] Angela Lashbrook, *AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind*, The Atlantic (Aug. 16, 2018), https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/

[15] *See* Jacob Snow, *supra* note 13.

[16] Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor (St. Martin's Press 2018).

[17] Arkansas: Colin Lecher, *What Happens When an Algorithm Cuts Your Health Care*, The Verge (Mar. 21, 2018), https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy/; Idaho: *Federal Court Rules Against Idaho Department of Health and Welfare in Medicaid Class Action*, ACLU (Mar. 30, 2016), https://www.aclu.org/press-releases/federal-court-rules-against-idaho-department-health-and-welfare-me

calculate how much home healthcare chronically ill Medicaid patients would receive. Due to an error, the system was significantly underprovisioning many people who required such care to survive. Patients were left to sit in their own waste, unable to access food when they were hungry, or to turn themselves to prevent bedsores.[18] If Legal Aid of Arkansas had not brought a case, and ultimately audited the system, it's possible that such harm would have persisted unchecked.[19]

It is not an overstatement to claim that the integration of AI is poised to reshape core domains. In education, we're seeing the implementation of facial recognition,[20] automated mental health monitoring,[21] and AI-based 'learning programs' that promise to track student progress and direct teachers.[22] These are often sold to school districts by vendors making claims about technical capabilities that, when we're able to examine them, frequently fall short. For example, a company called Gaggle licenses services to school districts that it claims increase safety and automatically detect students with mental health issues.[23] Gaggle's system requires blanket surveillance of student social media and private student communications sent using school networks. It uses AI to analyze the content of these communications, and to flag threats and identify students at risk. But on examination, it becomes clear that much of what gaggle detects are "minor violations," including use of profanity in private communications.[24] The company's claims that it helps students and improves safety are not firmly substantiated.

We are left with an understanding of Gaggle's sales pitch, but without answers to a number of socially significant questions that its technology raises: we don't know how a student might contest the company's interpretation of their communications, whether parents were fully informed before consenting to such surveillance, the way in which such monitoring may chill communication between peers, or how Gaggle's system might encode bias that could harm students with disabilities, students of color, or LGBTQ students (although we do know that in one school district the system flagged over a dozen students for communications that included

---

dicaid-class-action; Indiana: Alyssa Edes & Emma Bowman, *'Automating Inequality': Algorithms in Public Services Often Fail the Most Vulnerable*, All Things Considered (Feb. 19, 2018), https://www.npr.org/sections/alltechconsidered/2018/02/19/586387119/automating-inequality-algorithms-in-public-services-often-fail-the-most-vulnerab/.

[18] Caitlan Butler, *Changes to Medicaid Program Could Affect Union County Residents, Businesses*, El Dorado News-Times (Nov. 14, 2018), https://www.eldoradonews.com/news/2018/nov/14/changes-medicaid-program-could-affect-county/.

[19] *See* Colin Lecher, *supra* note 17.

[20] Mariella Moon, Facial recognition is coming to US schools, starting in New York, Engadget (May 30, 2019), https://www.engadget.com/2019/05/30/facial-recognition-us-schools-new-york/.

[21] Kaveh Waddell, *Schools Turn to AI to Monitor Students' Mental Health*, Axios (August 29, 2018), https://www.axios.com/school-filtering-for-self-harm-prevention-4bf481cc-a351-4f59-8cb3-318943548edc.html

[22] https://www.century.tech/

[23] https://www.gaggle.net/

[24] Benjamin Herold, *Schools Are Deploying Massive Digital Surveillance Systems. The Results Are Alarming*, Education Week (May 30, 2019), https://www.edweek.org/ew/articles/2019/05/30/schools-are-deploying-massive-digital-surveillance-systems.html

the word "gay").[25] Gaggle is one example among many in which AI is being applied in ways that have profound and life-changing implications without first ensuring that it is safe, helpful, unbiased, and doesn't pose risks to those it's meant to serve.

Corporate Control and Secrecy

AI technology affecting the lives and opportunities of hundreds of millions of people is being shaped by the incentives, values, and interests of a small handful of private companies.

AI is not a new set of technologies, and many of the core techniques that power AI systems, including neural nets, have been around for decades. The biggest changes recently have not been wholly new AI techniques (although we have seen improvements and innovations). What has changed drastically is the availability of massive amounts of data and vast computational resources. It's this that is behind the AI boom we see today. These are assets that only a handful of major tech companies have, and very few others do. This is one of the reasons why the US government contracts with companies like Amazon, instead of building its own infrastructure and AI. Without the legal protections afforded the private sector around privacy, and existing market reach and infrastructural economies of scale, it's virtually impossible to obtain the resources needed to create AI from scratch.

While there are many AI startups, most have significant budget allocated to licensing computational infrastructure from one of the big tech companies – namely Amazon, Microsoft, or Google. Most also struggle to acquire data, often scraping the web, buying data from data brokers, or signing deals in one or another domain (say, healthcare or education) to get access to relevant datasets. In short, the startup AI ecosystem we see today ultimately relies on the infrastructure provided by big companies, and thus the AI industry points back to a few large players.

This means that much of what we know about AI is written by the public relations and marketing departments of these same companies. They highlight benevolent uses and public benefit, often showcasing prototypes that haven't been validated beyond narrow test cases,[26] while remaining silent about the application of AI to fossil fuel extraction,[27] weapons development,[28] mass

---

[25] *See id.*
[26] Yossi Matias, *Keeping People Safe With AI-Enabled Flood Forecasting*, The Keyword (Sep. 24, 2018), https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/.
[27] Brian Merchant, *How Google, Microsoft, and Big Tech Are Automating the Climate Crisis*, Gizmodo (Feb. 21, 2019), https://gizmodo.com/how-google-microsoft-and-big-tech-are-automating-the-1832790799/.
[28] Jason Evangelho, *Microsoft Employees Upset About HoloLens As U.S. Military Weapon*, Forbes (Feb. 23, 2019), https://www.forbes.com/sites/jasonevangelho/2019/02/23/microsoft-employees-upset-about-hololens-as-u-s-military-weapon/#5e49f1524822/.

surveillance,[29] or the problems of bias and error.[30] We know about many of these more troubling cases thanks to researchers, investigative journalists, and whistleblowers.

When regulators, researchers, and the public seek to learn more, and to research and understand the potential harms of these systems, they are faced with structural barriers. The companies developing and deploying AI often exploit corporate secrecy laws, making testing, auditing, and monitoring extremely difficult, if not impossible.[31]

This structural secrecy, combined with overbroad laws used to threaten scientists, journalists, and algorithmic auditors,[32] makes it very difficult to validate company marketing promises. Access to fundamental information about AI systems, like where, how, and to what end they're being used, are classed as proprietary and confidential. Often even workers within these firms don't know where, and how, technology they contribute to will ultimately be applied.[33]

Lack of Accountability to Those Most at Risk

This pattern is particularly concerning given that those who will be most affected by such systems are rarely part of decision-making leading to the purchase and implementation of such systems, and often lack power to challenge subsequent uses. Tenants in rent-stabilized buildings in Brooklyn only learned of their landlord's plan to install a StoneLock brand facial recognition entry system after he made the decision to procure and install the system. A letter from New York State's Homes and Community Renewal (HCR), the agency overseeing rent-stabilized housing, informed tenants of the landlord's application to install the system. The tenants pushed back, citing well-documented problems of bias and inaccuracy, as well as the privacy concerns inherent in providing sensitive biometric data to a landlord with a history of punitive tenant surveillance.[34] Currently, the plan remains in limbo, with a lawyer representing the tenants noting that HCR does not have the jurisdiction or authority to adequately protect

---

[29] Drew Harwell, *Oregon Became a Testing Ground for Amazon's Facial-Recognition Policing. But What If Rekognition Gets It Wrong?*, Washington Post (Apr. 30, 2019), https://www.washingtonpost.com/technology/2019/04/30/amazons-facial-recognition-technology-is-supercharging-local-police/?utm_term=.50977aa76349/.

[30] Jeremy B. Merrill & Ariana Tobin, *Facebook Moves to Block Ad Transparency Tools — Including Ours*, ProPublica (Jan. 28, 2019), https://www.propublica.org/article/facebook-blocks-ad-transparency-tools/.

[31] Vinod Iyengar, *Why AI Consolidation Will Create the Worst Monopoly in US History*, TechCrunch (Aug. 24, 2016), https://techcrunch.com/2016/08/24/why-ai-consolidation-will-create-the-worst-monopoly-in-us-history/.

[32] *Sandvig v. Barr — Challenge to CFAA Prohibition on Uncovering Racial Discrimination Online*, ACLU (May 22, 2019), https://www.aclu.org/cases/sandvig-v-barr-challenge-cfaa-prohibition-uncovering-racial-discrimination-online/.

[33] Kate Conger and Cade Metz, *Tech Workers Now Want to Know: What Are We Building This For?*, N.Y. Times (Oct. 7, 2018), https://www.nytimes.com/2018/10/07/technology/tech-workers-ask-censorship-surveillance.html

[34] Rashida Richardson, *Letter in Support of Brooklyn Legal Services' Opposition Re: New York State Homes and Community Renewal Docket Nos. GS210005OD and GS210008OD*, AI Now Institute (May 1, 2019), https://ainowinstitute.org/dhcr-amici-letter-043019.pdf.

tenants, and encouraging the agency to reject the landlord's application given the lack of meaningful protections.[35]

Agreements between a vendor selling such a system on the one hand, and a business or institution with an interest in using it on the other, are most often reached behind closed doors. In the case of many large tech companies, the fact that such an agreement has been reached may itself be considered confidential. If the Brooklyn buildings had not been rent-stabilized, and thus had not required an application to a state agency to implement such changes, it's possible that the tenants would only have learned after a contract between the landlord and the vendor had been signed. In most cases, there's no requirement that such agreements take the interests of people who will be tracked, classified, and surveilled by AI systems into account. And since these systems are often integrated in ways that aren't visible – as part of larger infrastructures – those affected by them may not know that an AI system had a role in making a determination that impacted their life.[36]

Lack of Diversity and Its Implications

The technology affecting the lives and opportunities of billions of people is being shaped by the incentives, values, and interests of a small handful of private companies. And these companies are notoriously non-diverse. AI Now conducted a year-long study on the AI industry's diversity and inclusion practices, and the results are bleak. Women make up 10% of research staff at Google and just 15% at Facebook,[37] and the picture is worse when you examine available data on non-white workers. Only 2.5% of Google's employees are black, and only 3.6% are latinx, while Facebook and Microsoft are each at 4% for black workers and 6% for latinx workers.[38] We have no data on trans workers and other gender minorities, but anecdotal evidence is grim. This extends beyond industry. In academia, over 80% of AI professors are men, and only 18% of authors at leading AI conferences are women.[39] To give an example of how stark the diversity problem is, when BlackInAI co-founder, Timnit Gebru, first attended the major AI conference NeurIPS in 2016, she counted 6 black attendees out of 8,500 total.[40]

---

[35] *Brooklyn Tenants File Legal Opposition to Landlord's Application to Install Facial Recognition Entry System in Building*, Legal Services NYC (May 1, 2019), https://www.legalservicesnyc.org/news-and-events/press-room/1466-brooklyn-tenants-file-legal-opposition-to-landlords-application-to-install-facial-recognition-entry-system-in-building

[36] Shannon Liao, *Chinese Facial Recognition System Mistakes a Face on a Bus for a Jaywalker*, The Verge (Nov. 22, 2018), https://www.theverge.com/2018/11/22/18107885/china-facial-recognition-mistaken-jaywalker/.

[37] Tom Simonite, AI Is the Future—but Where Are the Women, Wired (Aug. 17, 2018), https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/.

[38] *Diversity & Inclusion*, Microsoft Global Diversity and Inclusion, https://www.microsoft.com/en-us/diversity/default.aspx ; Maxine Williams, *Facebook 2018 Diversity Report: Reflecting on Our Journey*, Facebook Newsroom (July 12, 2018), https://newsroom.fb.com/news/2018/07/diversity-report/.

[39] JF Gagne, *Global AI Talent Report 2019*, https://jfgagne.ai/talent-2019/

[40] Jackie Snow, *"We're in a Diversity Crisis": Cofounder of Black in AI on What's Poisoning Algorithms in Our Lives*, MIT Technology Review (Feb. 14, 2018),

The diversity crisis in the AI industry means that women, people of color, gender minorities, and other marginalized populations are excluded from contributing to the design of AI systems, from shaping how these systems function, and from determining what problems these systems are tasked with solving.[41] This influences which AI products get built, who they are designed to serve, and who benefits from their development. And in the case of AI, these determinations affect billions of people beyond company walls.

AI systems have evidenced a persistent pattern of gender and race-based discrimination (among other attributes and forms of identity), and in most cases, such bias mirrors and replicates existing structures of inequality.[42] To review a few examples: sentencing algorithms discriminate against black defendants;[43] chatbots easily adopt racist and misogynistic language when trained on online discourse;[44] and Uber's facial recognition doesn't work for trans drivers, locking them out of work.[45] Here we see a common theme: when these systems fail, they fail in ways that harm those who are already marginalized. I have yet to encounter an AI system that was biased against white men as a standalone identity.

Such bias can be the result of faulty training data, problems in how the system was designed or configured,[46] or bad or biased applications in real world contexts. In all cases it signals that the environments where a given system was created and envisioned didn't recognize or reflect on the contexts within which these systems would be deployed. Or, that those creating and maintaining these systems did not have the experience or background to understand the diverse environments and identities that would be impacted by a given system. Recent research from AI Now's Rashida Richardson shows vendors selling predictive policing systems failed to account for potential biases in the data such systems relied on, and thus to prevent harm.[47] In at least nine jurisdictions, predictive policing tools were being used or developed on data that was generated during periods where the police departments engaged in corrupt, racially biased, or

---

https://www.technologyreview.com/s/610192/were-in-a-diversity-crisis-black-in-ais-founder-on-whats-poisoning-the-algorithms-in-our/

[41] Kate Crawford, Artificial Intelligence's White Guy Problem, N.Y. Times (June 25, 2016), https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

[42] *See, e.g.*, Safiya Umoja Noble, Algorithms of Oppression: How Search Engines Reinforce Racism (NYU Press, 2013); Latonya Sweeney, *Discrimination in Online Ad Delivery*, 56 Comm. of the ACM 5, 44-45 (2013); Dorothy E. Roberts, *Book Review: Digitizing the Carceral State, 1696 Harv. L. Rev. 1695* (2019); Muhammad Ali et al., *Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Skewed Outcomes*, arXiv (Apr. 19, 2019), https://arxiv.org/pdf/1904.02095.pdf.

[43] Julia Angwin et al., *Machine Bias*, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[44] James Vincent, *Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less Than a Day*, The Verge (Mar. 24, 2016), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist/.

[45] Steven Melendez, *Uber Driver Troubles Raise Concerns About Transgender Face Recognition*, Fast Company (Aug. 9, 2018), https://www.fastcompany.com/90216258/uber-face-recognition-tool-has-locked-out-some-transgender-drivers/.

[46] *See* Colin Lecher, *supra* note 17.

[47] *See* Rashida Richardson et al., *supra* note 2.

unlawful policing practices and policies. Police data reflects the local practices, policies, and environment where it was collected, as there are no national standards on how police data is collected or used. This research showed that reliance on such "dirty data" increases the risk that these police technologies will reproduce and perpetuate past discrimination..This is another example of how the failure by vendors and developers to account for history and context risks encoding bias directly into their systems. And thus risk the lives and futures of those wrongfully profiled and targeted.

Beyond Technical Solutions

Both within the spaces where AI is being created, and in the logic of how AI systems are designed, the costs of bias, harassment, and discrimination are borne by the same people: gender minorities, people of color, and other under-represented groups. Similarly, the benefits of such systems – from profit to efficiency – accrue primarily to those already in positions of power, who again tend to be white, educated, and male.[48]

This points to a problem that goes beyond technical fixes for issues of AI bias and discrimination. AI systems are powerful. They allow those who use them to profoundly influence people's lives, across multiple domains. And the ability to access and use these systems is not evenly distributed. Given the expense of creating AI, the cost of maintaining these systems, and the market incentives driving their development, it is almost always those who already have power who are in the position of applying AI systems, often on those who don't. Immigration and Customs Enforcement (ICE) had used a risk assessment algorithm at the border since 2013, meant to help determine whether an immigrant should be detailed or released on bond. An academic study examining the algorithm found that it had been modified a number of times in an attempt to mitigate the overly-punitive behavior of ICE agents,[49] but this technical bias fix failed to solve the problem. Then, in 2017 ICE modified the algorithm again, to *only* suggest detain, and never release. The impact of this change was profound. ICE detained over 43,000 immigrants with no criminal history in 2017, more than three times as many as the previous year.[50]

Here we see the way in which the problems of AI go beyond bias and inaccuracy at the technical level, to fundamental issues of power and control. This means our approach to these problems must expand beyond the technical, asking who is harmed by these systems? Who benefits? And who gets to decide?

---

[48] *See* Sarah Myers West et al., *supra* note 2.
[49] Robert Koulish & Ernesto F. Calvo, *The Human Factor: Algorithms, Dissenters, and Detention in Immigration Enforcement*, ILCSS WORKING PAPER | NO. 1 (March 16, 2019), *available at* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3355663.
[50] Mica Rosenberg & Reade Levinson, *Trump's Catch-And-Detain Policy Snares Many Who Have Long Called U.S. Home*, REUTERS (June 20, 2018), https://www.reuters.com/investigates/special-report/usa-immigration-court/

Ethics Is Not Enough

In the face of evidence of bias, error, and misuse of AI systems, we have seen tech companies and the AI field beyond adopt ethical principles and guidelines, and form ethical oversight committees and boards. It is encouraging that those creating such powerful technologies are recognizing their potential harms and moving to address these.

However, these codes, guidelines and ethics boards do not substitute for meaningful accountability and oversight. To date we have no examples where such ethical proclamations are backed by public enforcement mechanisms, nor any that attach clear consequences for failing to live up to ethical ideals. While the public may be able to compare a decision already made by an AI company to its guiding principles, this allows no insight into decision-making, nor does it give anyone outside of the company the power to reverse or guide such a decision.

Such a turn to ethics may also serve to stall movements toward more substantive regulation and public accountability measures, working to deflect criticism by acknowledging that problems exist, without ceding any power to regulate or transform the way technology is developed and applied.[51]

The case of Google is instructive here. Among the list of applications that Google promises not to pursue as a part of its AI Principles are "technologies whose purpose contravenes widely accepted principles of international law and human rights."[52] This commitment was tested in 2018 when whistleblowers revealed that the company was quietly developing a censored version of its search engine (which relies extensively on AI capabilities) for the Chinese market, code-named Dragonfly.[53] Human rights organizations condemned the project as a violation of human rights law, and as such, a violation of Google's AI principles.[54] [55] While there are indications that the project has been cancelled following Google worker organizing and public outcry, its continued development in the face of the company's principles was never explained,

---

[51] Ben Wagner, *Ethics as Escape From Regulation: From Ethics-Washing to Ethics-Shopping?*, in Being Profiled: Cogitas Ergo Sum, ed. Mireille Hildebrandt. (Amsterdam University Press, forthcoming 2019), https://www.privacylab.at/wp-content/uploads/2018/07/Ben_Wagner_Ethics-as-an-Escape-from-Regulation_2018_BW9.pdf.

[52] Sundar Pichai, *AI at Google: Our Principles*, The Keyword, June 7, 2018, https://www.blog.google/technology/ai/ai-principles/.

[53] Ryan Gallagher, *Google Plans to Launch Censored Search Engine in China, Leaked Documents Reveal*, The Intercept, August 1, 2018, https://theintercept.com/2018/08/01/google-china-search-engine-censorship/.

[54] Ronald Deibert, Rebecca Mackinnon, Xiao Qiang, and Lokman Tsui, *Open Letter to Google on Reported Plans to Launch a Censored Search Engine in China*, Amnesty International (August 28, 2018), https://www.amnesty.org/en/documents/document/?indexNumber=ASA17%2f9001%2f2018.

[55] Kate Conger and Daisuke Wakabayashi, *Google Employees Protest Secret Work on Censored Search Engine for China*, N.Y. Times, September 10, 2018, https://www.nytimes.com/2018/08/16/technology/google-employees-protest-search-censored-china.html.

and the principles themselves seem to have had little impact on directing the company to reconsider the effort.[56]

Organizing and Accountability

Over the past year, we have seen mounting pushback, as organizing and resistance to the development and application of AI systems has grown. Amazon Warehouse workers in Minneapolis protested Amazon's automated management system, pushing back against a system of worker control that continually extracted more labor at the expense of workers' health and well-being.[57] Similarly, Uber drivers organized a nationwide strike, protesting the centralized, algorithmic platform that arbitrarily cuts driver wages without recourse.[58] And even students at a school where Facebook-backed "personalized" education systems were used staged protests, demanding less dehumanizing forms of education.[59]

We've also seen pushback inside tech companies. Workers across the industry have organized and protested unethical and secretive projects,[60] [61] [62] [63] [64] demanding to have a say in what they

---

[56] Mark Bergen, *Google CEO Tells Staff China Plans Are 'Exploratory' After Backlash*, Bloomberg (August 17, 2018), https://www.bloomberg.com/news/articles/2018-08-17/google-ceo-is-said-to-tell-staff-china-plans-are-exploratory

[57] Chavie Lieber, *Muslim Amazon Workers Say They Don't Have Enough Time to Pray. Now They're Fighting for Their Rights*, Vox (Dec. 17, 2018), https://www.vox.com/the-goods/2018/12/14/18141291/amazon-fulfillment-center-east-africa-workers-minneapolis

[58] Faiz Siddiqui, *Uber and Lyft Drivers Strike for Pay Transparency — After Algorithms Made It Harder to Understand*, Washington Post (May 8, 2019), https://www.washingtonpost.com/technology/2019/05/08/uber-lyft-drivers-strike-pay-transparency-after-algorithms-made-it-harder-understand/?utm_term=.b4c939121e0e

[59] Susan Edelman, *Brooklyn Students Hold Walkout in Protest of Facebook-Designed Online Program*, N.Y. Post (Nov. 10, 2018), https://nypost.com/2018/11/10/brooklyn-students-hold-walkout-in-protest-of-facebook-designed-online-program/

[60] Dave Lee, *Microsoft Staff: Do Not Use HoloLens for War*, BBC News (Feb. 22, 2019), https://www.bbc.com/news/technology-47339774/.

[61] Ryan Gallagher, *Google Dragonfly*, The Intercept (Aug. 1, 2018 – May 14, 2019), https://theintercept.com/collections/google-dragonfly-china/ (collecting 25 articles of Google Dragonfly coverage).

[62] Kate Conger, *Amazon Workers Demand Jeff Bezos Cancel Face Recognition Contracts With Law Enforcement*, Gizmodo (June 21, 2018), https://gizmodo.com/amazon-workers-demand-jeff-bezos-cancel-face-recognitio-1827037509

[63] Scott Shane and Daisuke Wakabayashi, *'The Business of War': Google Employees Protest Work for the Pentagon*, N.Y. Times (April 4, 2018), https://www.nytimes.com/2018/04/04/technology/google-letter-ceo-pentagon-project.html/.

[64] Becky Peterson, *Salesforce Employees Are Upset Over the Company's Work With U.S. Customs and Border Protection as Silicon Valley Grapples With the Government's Use of Tech*, Business Insider Prime (Jun 25, 2018), https://www.businessinsider.com/salesforce-employees-protest-work-customs-border-protection-agency-2018-6/.

build,[65] and connecting AI's ethical issues with problems of workplace discrimination, harassment, and abuse.[66]

Currently, such organizing and activism comprises one of the primary modes of accountability working to check the biased and oppressive deployment of AI technologies, and to improve diversity and equity across the industry.[67] [68] [69] [70] I have proudly organized with fellow tech workers at Google and beyond, recognizing the urgent need for such accountability.[71] What is clear to those of us who have been researching these issues, and those of us familiar with the tech industry, is that we need much more. Accountability, transparency, oversight, and measures that ensure those most at risk of harm are at the heart of AI decision making.[72] [73]

The Path Forward: AI Now's Policy Recommendations for Congress:

Congress has a window to act, and the time is now. Powerful AI systems are currently reshaping our lives and social institutions in ways that we aren't able to measure or contest. These systems are developed and deployed by a handful of private companies whose market interests don't always align with the public good, and who shield these systems from accountability behind claims of corporate secrecy. When we are able to examine these systems, too often we find that they are biased and inaccurate in ways that replicate historical patterns of discrimination.

It is imperative that lawmakers regulate to ensure that these systems are accountable, accurate, contestable, and that those most at risk of harm have a say in how, and whether, they are used. As first steps toward this goal, AI Now recommends that lawmakers:

---

[65] *See* Kate Conger and Daisuke Wakabayashi, *supra* note 55. https://www.nytimes.com/2018/08/16/technology/google-employees-protest-search-censored-china.html/.
[66] Google Walkout for Real Change, *20,000 Google Employees and Contractors Participate in Global "Walkout For Real Change"* (Nov 2, 2018), https://medium.com/@GoogleWalkout/google-employees-and-contractors-participate-in-global-walkout-for-real-change-389c65517843/.
[67] Daisuke Wakabayashi, *Google Ends Forced Arbitration for All Employee Disputes*, N.Y. Times (Feb. 21, 2019), https://www.nytimes.com/2019/02/21/technology/google-forced-arbitration.html/.
[68] BBC News, *Microsoft Responds to Female Harassment Claims*, BBC (April 5, 2019), https://www.bbc.com/news/technology-47826249/.
[69] Nick Bastone, *Google Is Likely to End Its Efforts to Build a Censored Search Engine for China, Says Report*, Business Insider (Dec. 17, 2018), https://www.businessinsider.com/google-dragonfly-china-canceled-2018-12/.
[70] Daisuke Wakabayashi & Scott Shane, *Google Will Not Renew Pentagon Contract that Upset Employees*, N.Y. Times (June 1, 2018), https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html/.
[71] Kyle Wiggers, *How Google Treats Meredith Whittaker Is Important to Potential AI Whistleblowers*, VentureBeat (April 24, 2019), https://venturebeat.com/2019/04/24/how-google-treats-meredith-whittaker-is-important-to-potential-ai-whistleblowers/.
[72] Frank Pasquale, Odd Numbers: Algorithms Alone Can't Meaningfully Hold Other Algorithms Accountable, Real Life Mag (Aug. 20, 2018), https://reallifemag.com/odd-numbers/.
[73] *See* Meredith Whittaker et al., *supra* note 2.

1. **Require Algorithmic Impact Assessments in both Public and Private Sectors before AI Systems are Acquired and Used**

   In 2018, AI Now published an Algorithmic Impact Assessment (AIA) framework, which offers a means for assessing algorithmic systems in government, including AI based systems, and providing the public with transparency and decision-making power.[74] AIAs draw directly from impact assessment frameworks in environmental protection, human rights, privacy, and data protection policy domains by combining public agency review and public input.[75] When implemented in government, AIAs provide both the agency and the public the opportunity to evaluate the potential impacts of the adoption of an algorithmic system before the agency has committed to its use. AIAs also require ongoing monitoring and review, recognizing the dynamic contexts within which such systems are applied.

   The framework has been adopted in Canada, and is being considered by local, state, and national governments globally.[76] Though it was originally proposed to address concerns associated with government use of automated decision-making systems, the framework should also be mandated for private companies, ensuring review and public engagement before a product or service is used in ways that impact the public. This can provide companies opportunities to assess and possibly mitigate adverse or unanticipated outcomes during the development process. It also provides the government and public with greater transparency, along with a means to strengthen existing consumer accountability mechanisms. By requiring that proposed technologies be reviewed by the communities who will be most affected by their use, AIAs work to empower those most at risk, and to encourage the development of safer and more ethical technologies.

2. **Require Technology Companies to Waive Trade Secrecy and Other Legal Claims That Hinder Oversight and Accountability Mechanisms**

   Corporate secrecy laws are a barrier to oversight, accountability, and due process when they are used to obscure technologies used in ways that affect the public. They can

---

[74] *See* AI Now Inst., Algorithmic Accountability Policy Toolkit, *supra* note 2.
[75] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Calif. L. Rev. 671 (2016).
[76] European Parliament Panel for the Future of Sci. and Tech., A Governance Framework for Algorithmic Accountability and Transparency: Study (Apr. 4, 2019), http://www.europarl.europa.eu/stoa/en/document/EPRS_STU(2019)624262; Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong., (1st Sess. 2019), https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019%20Bill%20Text.pdf; Canada Treasury Board's Directed Automated Decision-Making, Autonomisation des Acteurs Judiciaires par la Cyberjustice (Nov. 25, 2018), https://www.ajcact.org/2018/11/25/canada-treasury-boards-directive-on-automated-decision-making/

inhibit necessary government oversight and enforcement of consumer protection laws,[77] which contribute to the "black box effect," making it hard to assess bias, contest decisions, or remedy errors. Anyone procuring these technologies for use in the public sector should have the right to demand vendors waive these claims before entering into any agreements. Additionally, limiting the use of these legal claims across the board will help facilitate better oversight by state and federal consumer protection agencies and enforcement of false and deceptive practice laws.

3. **Require Public Disclosure of Technologies That Are Involved in Any Decisions About Consumers by Name and Vendor/s**

Those most affected by AI's use in sensitive social domains often don't know that a given system was used, or where and how it might have shaped their lives. Consumers deserve to know about which AI systems are making decisions about them, or affecting the types of services, resources, or opportunities made available to them. Requiring disclosure of which AI systems are used, in what context, along with which companies developed such systems, will provide consumers with the notice necessary to enforce their due process rights. We need to ensure meaningful insight and transparency. This is especially urgent given that infrastructure owned by the major technology companies is often licensed and used by other businesses. Large technology companies license AI application program interfaces (APIs), or "AI as a service" to third parties, who apply them to one or another purpose.[78] These business relationships, in which one organization repurposes potentially flawed and biased AI systems created by large technology companies, are rarely disclosed to the public, and even the fact that there is such a relationship is often protected under nondisclosure agreements. It is hard, if not impossible, to simply know if an AI model is being used in a given context, let alone to know that such a model was created by Facebook, Google, or Amazon. Thus, understanding the implications of bad, biased, or misused models is effectively impossible, especially for those most at risk.

4. **Enhance Whistleblower Protections and Protections for Conscientious Objectors within Technology Companies**

Organizing and resistance by tech workers has emerged as a force for accountability

---

[77] *Houston Fed'n of Teachers, Local 2415 v. Houston Indep. Sch. Dist.*, 251 F.Supp.3d 1168 (S.D. Tex. 2017).

[78] *Cognitive Services*, Microsoft Azure, https://azure.microsoft.com/en-us/services/cognitive-services/ (last visited June 16, 2019); *AI and Machine Learning Products*, Google Cloud, https://cloud.google.com/products/ai/ (last visited June 16, 2019); *Tools for Advancing the World's AI*, Facebook Artificial Intelligence, https://ai.facebook.com/tools/ (last visited June 16, 2019); Machine Learning on Amazon AWS, Amazon Web Services, https://aws.amazon.com/machine-learning/ (last visited June 16, 2019); Matt Murphy & Steve Sloane, *The Rise of APIs*, TechCrunch (May 21, 2016), https://techcrunch.com/2016/05/21/the-rise-of-apis/.

and ethical oversight.[79] Alongside organizers, whistleblowers have provided a crucial public benefit, revealing products and problems that may not otherwise be visible to relevant oversight bodies, or even to the majority of workers within a given company. Whistleblowers in the technology industry can be a crucial component to government oversight, serving the public interest by revealing troubling and unethical initiatives poised to affect millions of people. In light of their service, and the critical role they are playing, they should be provided enhanced protections.

[79] Daisuke Wakabayashi & Scott Shane, *Google Will Not Renew Pentagon Contract that Upset Employees*, N.Y. TIMES (June 1, 2018), https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html; Avie Schneider, *Microsoft Workers Protest Army Contract With Tech 'Designed to Help People Kill'*, NPR (Feb. 22, 2019), https://www.npr.org/2019/02/22/697110641/microsoft-workers-protest-army-contract-with-tech-designed-to-help-people-kill; Mark Bergen & Nico Grant, *Salesforce Staff Ask CEO to Revisit Ties with Border Agency*, BLOOMBERG (June 25, 2018), https://www.bloomberg.com/news/articles/2018-06-25/salesforce-employees-ask-ceo-to-revisit-ties-with-border-agency.

Meredith Whittaker Biography

Meredith Whittaker is a Distinguished Research Scientist at New York University, Co-founder and Co-director of the AI Now Institute, and the founder of Google's Open Research group. She has over a decade of experience working in industry, leading product and engineering teams. She co-founded M-Lab, a globally distributed network measurement system that provides the world's largest source of open data on internet performance. She has also worked extensively on issues of data validation and privacy. She has advised the White House, the FCC, the City of New York, the European Parliament, and many other governments and civil society organizations on artificial intelligence, internet policy, measurement, privacy, and security. She is the co-founder and co-director of the AI Now Institute at NYU, which is a leading university institute dedicated to researching the social implications of artificial intelligence and related technologies.

Chairwoman JOHNSON. Thank you. Mr. Jack Clark.

TESTIMONY OF JACK CLARK,
POLICY DIRECTOR, OPENAI

Mr. CLARK. Chairwoman Johnson, Ranking Member Baird, and Committee Members, thank you for inviting me today. I'm the Policy Director for OpenAI, a technical research lab based in San Francisco.

I think the reason why we're here is that AI systems have become—and I'm using air quotes—good enough to be deployed widely in society, but lots of the problems that we're going to be talking about are because of "good enough" AI. We should ask, "good enough for who?", and we should also ask "good enough at what?"

So to give you some context, recent advances in AI have let us write software that can interpret the contents of an image, understand wave forms in audio, or classify movements in video, and more. At the same time, we're seeing the resources applied to AI development grow significantly. One analysis performed by OpenAI found that the amount of computing power used to train certain AI systems had increased by more than 300,000 times in the last 6 years, correlating to significant economic investments on the part of primarily industry in developing these systems.

But though these systems have become better at doing the tasks we set for them, they display problems in deployment. And these problems are typically a consequence of people failing to give the systems the right objectives or giving them the right training data. Some of these problems include popular image recognition systems that have been shown to accurately classify products from rich countries and fail to classify products from poor countries, voice recognition systems that perform extremely badly when dealing with people who are speaking in English that is heavily accented, or commercially available facial recognition systems that consistently misclassify or fail to classify people with darker skin tones.

So why these issues arise is because many modern machine learning systems automate tasks that require people to make value judgments. And so when people make value judgments, they encode their values into the system, whether that's the value of who's got to be in the dataset or what the task is that it's solving. And because, as my co-panelists have mentioned, these people are not from a particularly diverse background, you can also expect problems to come from these people selecting values that apply to many people.

These systems can also fail as a consequence of technical issues, so image classification systems can be tricked using things known as adversarial examples to consistently misclassify things they see in an image. More confusingly and worryingly, we found that you can break these systems simply by putting something in an image that they don't expect to see. And one memorable study did this by placing an elephant in a room, which would cause the image recognition system to misclassify other things in that room even though it wasn't being asked to look at it. So that gives you a sense of how brittle these systems can be if they're applied in the context which they don't expect.

I think, though, that these technical issues are in a sense going to be easier to deal with than the social issues. The questions of how these systems are deployed, who is deploying them, and who they're being deployed to help or surveil are the questions that I think we should focus on here. And to that end I have a few suggestions for things that I think government, industry, and academia can do to increase the safety of these systems.

First, I think we need additional transparency. And what I mean by transparency is government should convene academia and industry to create better tools and tests and assessment schemes such as the, you know, algorithmic impact assessment or work like adding a label to datasets which are widely used so that people know what they're using and have tools to evaluate their performance.

Second, government should invest in its own measurement assessments and benchmarking schemes potentially by agencies such as NIST. The reason we should do this is that, as we develop these systems for assessing things like bias, we would probably want to roll them into the civil sector and have a government agency perform regular testing in partnership with academia to give the American people a sense of what these systems are good at, what they're bad at, and, most crucially, who they're failing.

Finally, I think government should increase funding for interdisciplinary research, a common problem is these systems are developed by a small number of people from homogenous backgrounds, and they can also be studied in this way because grants are not particularly friendly to large-scale interdisciplinary research projects. So we should think about ways we can study AI that brings together computer scientists, lawyers, social scientists, philosophers, security experts, and more, not just 20 computer science professionals and a single lawyer, which is some people's definition of interdisciplinary research.

So, in conclusion, I think we have a huge amount of work to do, but I think that there's real work that can be done today that can let us develop better systems for oversight and awareness of this technology. Thank you very much.

[The prepared statement of Mr. Clark follows:]

Written Testimony of Jack Clark
Policy Director
OpenAI

HEARING ON

"Artificial Intelligence: Societal and Ethical Implications'"

BEFORE THE

House Committee on Science, Space, & Technology

June 26th, 2019.

1: Introduction.

Chairwoman Johnson, Ranking Member Lucas, and committee members, thank you for the opportunity to testify about this critical subject. This hearing is informed by my work at OpenAI, an artificial intelligence research and development company seeking to build general-purpose AI systems to benefit all of humanity. It is also informed by my work as a member of the Steering Committee for the AI Index, a Stanford initiative to track, measure, and analyse the progress and impact of AI technology.

When thinking about the ethical and societal challenges of AI, we must remember AI is a product of the environment it is developed in, and it reflects the inherent biases of the people and institutions that built it. Therefore, when we think about how AI interacts with society, we should view it as a *social system* rather than a technological system, and this view should guide the sorts of policies we consider when thinking about how to govern it.

For the purposes of this hearing I will discuss a relatively narrow subset of AI: recent advances in machine learning, oriented around pattern recognition. Some of these techniques are relatively immature, but have recently become 'good enough' for various deployment use cases. Crucially, 'good enough' isn't the same as 'ideal', and 'good enough' systems exhibit a range of problems and negative externalities which should require careful thinking during deployment. And whenever a system is "good enough" we should ask "for who?".

For this testimony, I will:

- Briefly outline recent progress in the field of artificial intelligence.
- Outline some of the ways in which contemporary and in-development systems can fail.
- Discuss the tools we have today to deal with such failures.
- Outline how government, industry, and academia can collectively address concerns around the development and deployment of AI systems.

1.1: Why we're here: We've entered the era of "good enough" AI

There are two classes of systems which are predominantly deployed today[1] - systems that classify the world according to an objective defined by a human, and things that predict something about the world and take an action. (As this hearing is predominantly focused on systems being deployed today or likely to be deployed in the future, I am limiting my overview here to the bits of AI which are gaining the most commercial interest.)

For classification, we have recently figured out how to create AI systems that can crudely mimic the capabilities of a couple of human senses: specifically, vision and hearing. By this, I mean

[1] Note that this description avoids discussion of 'expert systems' and other AI approaches which have been developed in prior decades and which have been deployed in parts of society since the 1980s. The focus of this testimony is on machine learning systems and specifically ones that primarily use deep learning - that's because these systems have broad capabilities and are being broadly deployed.

that recent advances in the field of 'machine learning' have let us develop systems that can - given a large enough dataset and computational power - learn to map labels to information extracted from images and audio. For instance, systems that assign a label to an image, or a part of one, like labeling fruit as being safe or rotten in farming, or a social network platform correctly identifying an individual in a photo, or a surveillance system classifying the actions of people in public spaces like train stations to identify suspicious activities.

To give a sense of the underlying pace of progress for this capability, we can look at the results of the 'ImageNet' object recognition competition: in 2010 a computer could be shown an image and, about **72%** of the time, come up with a list of five labels for the image, of which one would be correct. By 2017, this accuracy had climbed to above **97%**[2] - and progress is continuing.[3] This is progress on a particular dataset, but it relates to larger technological advancements, which loosely correlate to better performance on other specific tasks, like analyzing security camera footage, or spotting animals in nature. Similarly, for the field of speech recognition - that is, accurately transcribing speech - performance on one major benchmark has increased from **84%** in 2011 to **95%** in 2017[4].

However, these capabilities can degrade when exposed to things they haven't been trained on, like people of demographics different to the underlying dataset, or even products popular in "poor" countries.[5]

Meanwhile, research in reinforcement learning[6] has driven advancements in systems that can learn to act autonomously in specific circumstances. These systems can display their own patterns of failure, but it should be noted they are predominantly being research today, rather than widely deployed. (You can track the evolution of the capabilities of research systems here by looking at the complexity of the environment the agent can achieve an objective within. So, what does that look like? In 2013 we could use these systems to learn to play old Atari games like *Breakout!* and *Space Invaders*, in 2016 we could use such systems to beat humans at complex board games like *Go*, and in 2018 we could use these systems to compete with humans in very complex, real-time strategy video games like *StarCraft II* and *Dota 2*.)

The progress in these domains is impressive and worthy of attention, because they roughly correlate to contemporary or future societal impacts of AI: these performance increases, and associated ones in other domains, have led many AI systems to go from 'barely usable' to 'good

---

[2] Some research indicates that this exceeds human performance at this task. For more, see Andrej Karpathy "What I learned from competing against a ConvNet on ImageNet" http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/
[3] AI Index 2018 report, page 47. For more, see: https://aiindex.org
[4] AI Index 2017 report, page 31. For more, see: https://aiindex.org/2017/
[5] Does Object Recognition Work For Everyone?, DeVries et al: https://arxiv.org/abs/1906.02659
[6] Reinforcement learning is where you have an AI agent and a simulator (for instance, a flight simulator); you give the AI an objective (e.g, fly the plane from here to Spain), and then you have an AI system try to achieve this goal. The AI system will fail a lot, and each time it fails you restart the simulator and it tries again - eventually, the system will learn how to fly the plane to achieve the objective.

enough' in terms of real world deployment[7]. Given that AI is also a social system and has recently attained 'good enough' performance, we should ask good enough *for whom?* and good enough *at what?*

As my other panellists for this testimony will show, these systems, when deployed, frequently exhibit biases, and these biases can manifest as *inequitable access to the benefits of AI*[8] or *false positive identification by AI systems*. They also exhibit problems related to the process surrounding the design and deployment of AI systems, and some longer-term issues with the learning algorithms used to implement some AI systems.

We can expect progress in AI from both a research and a deployment view to continue, because of:

- Massive increases in the numbers of students involved in academic AI programs across the world.
- New funding from a variety of governments[9] and industry
- Falling costs of both computers and data storage systems.
- Ongoing algorithmic improvements.
- Commercial pressures; now that AI is "good enough" it makes economic sense for a large number of actors to invest in its development.

1.2: AI progress and economic incentives

Last year, OpenAI carried out an analysis in which we reviewed research papers relating to AI that were published in the last few years and analyzed the total amount of computational resources dedicated to the development of such systems. Our analysis showed that this amount had increased by **300,000X** over the past six years. The systems which fit this trend spanned use cases from image recognition, to machine translation, to strategic game playing systems. This trend correlates both to the increasing capabilities of some of these systems, and the increasing economic expenditures of large AI research and development organizations. (To put 300,000X in perspective, Moore's Law - that is, the 70-year trend that computers tend to double in capability every 18 months, would generate a **12X** increase over this same period.)

Many recent breakthroughs in AI systems for purposes like image recognition, speech recognition, machine translation, game playing, are correlated with this increasing compute

---

[7] If we were to define a turning point in this domain it might be around 2017 - that's when Google# (a subsidiary of Alphabet Inc.) described itself as an 'AI first' company, and other large companies signalled larger commitments to AI.

[8] For instance, research has shown that commercially deployed image AI systems from companies such as Amazon and others have significantly higher error rates at classifying females with darker skin tones. See: Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products, by Inioluwa Deborah Raji and Joy Buolamwini. http://www.aies-conference.com/wp-content/uploads/2019/01/AIES-19_paper_223.pdf

[9] Including, I hope, additional funding from the US government.

usage trend. This correlates to increasing economic expenditures by the companies deploying or researching the systems. This number also implies significant spending by people on the underlying computational systems required to train these AI systems, so the long-term trend could be altered by larger economic or R&D forces.

This number implies two things:

- 1) AI may progress more rapidly than peoples' intuitions would suggest, as people are bad at modelling what 300,000X increases correlate to.
- 2) We can expect the technical weaknesses of AI systems to 'scale up' with the amount of computational power poured into them, unless we develop smarter algorithms and better systems of governance for the organizations that develop them. This means that the ways AI algorithms fail at small scale can potentially be amplified and cause more harm when these failures occur in larger systems.

2. When 'good enough' AI goes bad.

I think there are two broad but related classes of failure we should think about here: when an AI system fails as a consequence of the *process* humans use when developing the system, and when an AI system fails as a consequence of the *learning algorithm* it has been equipped with[10]. For the purposes of this hearing, I think that failures of process are currently more numerous and consequential for society, while failures of algorithms may be significant in the long-term but are not as commonly seen in the wild today.

2.1 Process failures

Process failures typically manifest as an AI system failing dramatically during deployment, usually as a consequence of it being surprised by something. Unlike humans, AI systems are terrible at adapting to surprising situations, so these failures are typically severe as they speak to an underlying deficiency in the system. The system is typically surprised by something because it hasn't been built in a way that fully appreciates the context of the environment it is being deployed in.

Here are some examples of ways in which either researched or deployed systems have failed:

- Google's 'Google Photos' application incorrectly classified a black male as a gorilla. This failure was likely a consequence of the company not gathering enough data to teach its

[10] For a fuller overview of the various ways AI systems can fail - including systems currently on the frontier of AI research - please refer to "Concrete Problems in AI Safety" by Amodei et al (2016) https://arxiv.org/abs/1606.06565 and "Building safe artificial intelligence: specification, robustness, and assurance" by DeepMind Safety Research (https://medium.com/@deepmindsafetyresearch/building-safe-artificial-intelligence-52f5f75058f1)

systems to consistently characterise black males and gorillas accurately, and not having sufficient testing regimes to identify this issue prior to deployment.

- IBM's 'Watson' healthcare system would sometimes recommend "unsafe and incorrect" cancer treatments, according to a report by STAT News, with the flaws emanating from improper dataset selection and improper processes for collecting people's opinions about what effective treatments were[11].

2.2 Learning algorithm failures

A good way to think about artificial intelligence systems and failure is that when they go wrong, it is usually because they achieved the specification but not the spirit of the described rule; these actions can frequently seem inappropriate or unsafe to a human. Sometimes these problems relate to weaknesses in the algorithm used itself, and other times they relate to humans mis-specifying the objectives of the algorithm.

- **Brittleness:** Image recognition systems can fail as a consequence of imperceptible variations in the appearance of digital and real images - images that cause them to fail are known within machine learning as 'adversarial examples'[12]. They can also fail as a consequence of dealing with unanticipated things - in one memorable example, researchers showed that by superimposing an image of an elephant onto an otherwise normal image, they could reliably cause image recognition systems to fail to classify other parts of the image[13].

- **Mis-specified rewards:** When training an AI system to complete a boat race in a video game, OpenAI gave the system the objective of getting as many points as possible, after observing that points typically correlated to winning the race. Our boat found a bug in the game that meant it could get a high score by navigating to a lagoon in the center of the race and spinning itself around to repeatedly hit various high scoring items, while setting itself on fire[14].

- **Mis-specified rewards:** When training a simulated robot to move its arm to move a hockey puck from one point of a table to another, o=OpenAI's robot instead learned to move the entire table to move the puck, rather than sliding it deftly, as we had intended. This would be dangerous in a real-world setting, and even if you installed safety systems

[11] For more, please refer to IBM's Watson Supercomputer recommended "unsafe and incorrect" cancer treatments, internal documents show, by Casey Ross for Stat News (2018). https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/
[12] For more, see 'Explaining and Harnessing Adversarial Examples' by Goodfellow et al, (2014). https://arxiv.org/abs/1412.6572
[13] For more, see The Elephant in the Room by Rosenfeld et al, (2018). https://arxiv.org/abs/1808.03305
[14] For more information, please refer to: https://openai.com/blog/faulty-reward-functions/

on the robot the fact this happens indicates other unanticipated behaviors could occur during training of the AI system.

- **Unexpected exploits:** Other examples abound, and are regularly collected and analyzed by the AI community. For example: A four-legged evolved agent trained to carry a ball on its back discovered that it could drop a ball into a leg joint and then wiggle across the floor without the ball ever dropping; in another scenario an agent chose to kill itself at the end of the first level of a game so it could avoid losing in level 2 of the game, and so on[15].

2.3 Process + Learning Algorithm Failures

Many failures occur as a consequence of process failures as well as learning algorithm failures. I think these situations are where many of the hardest problems occur, because they typically require a combination of technical and social analysis to understand and respond to. Some examples of failures of these types could include:

- Recommendation engines: Today, many companies around the world are seeking to use machine learning to learn to recommend products or services to people. When these systems fail it's usually a consequence of the underlying learning algorithm achieving a mis-specified objective (for instance, optimizing for engagement when showing people videos, which can lead to people consuming more videos that they find engaging, which can sometimes correlate to extremist content[16]), as well as the organization not doing enough direct study of the end effects on its users.

3. What can academia, government, and industry do to address these issues?

Technological fixes alone will be insufficient to address potential impacts of these technologies - this work will require careful coordination between industry, academia, and government during the development and deployment of these systems. However, a list of work without the accompanying resources to carry it out is useless, so I feel it is prudent for the government to consider increasing its own ability to measure, analyze, benchmark, and forecast the

---

[15] For many more examples, please refer to: https://vkrakovna.wordpress.com/2018/04/02/specification-gaming-examples-in-ai/

[16] YouTube recently announced plans to remove thousands of extremist videos located on the web video service, according to *The New York Times* (June, 2019). https://www.nytimes.com/2019/06/05/business/youtube-remove-extremist-videos.html

development and application of AI systems, and to increase the funding it assigns to AI development[17] so academia is better equipped to solve these issues.

3.1: Government interventions

I think government has a profoundly important role to play here, chiefly by funding initiatives to gather more information about the progress and impact of AI systems. I believe it can step into this role via modest investment in its own capabilities to measure, assess, and forecast aspects of AI progress and impact. We need the equivalent of a publicly funded weather forecasting service for the ways in which AI is evolving so that we can better orient ourselves with regard to contemporary opportunities and problems and better spot problems and solutions that are over the horizon.

Specifically, I think government should intervene in the following ways:

- **Measurement, assessment, and analysis of deployed systems:** It would be helpful for the government to continuously benchmark for-sale or deployed machine learning systems for societally harmful failures, such as bias. Today, numerous academic researchers have developed datasets that deployed systems can be tested against; and we should consider building a 'bias test suite'[18], which government - potentially via agencies such as NIST - can develop as a resource for industry and academia.

- **Transparency in government AI procurement:** Today, it's difficult to get a sense for what AI systems are deployed[19]. Government can make a difference here by increasing the transparency with which federal agencies procure and deploy AI systems. This would equip academia with more information to use to study the impact of such systems, and would help further our knowledge about what responsible development and deployment of these systems looks like.

- **Funding:** We should increase the funding we allocate to artificial intelligence research and development in academia, while also increasing the resources to government agencies that can help coordinate actions between industry, government, and academia. I think that some existing proposed legislation, such as The Artificial Intelligence Initiative Act, could be helpful here. This legislation proposes increased funding for NIST, which would help that agency conduct more measurement and assessment of AI systems,

[17] This should be net-new funding for scientific research, rather than funding that detracts from existing research initiatives.
[18] Such a suite could consist of multiple datasets which systems can be tested against to show equitable effectiveness across a diverse set of people and objects.
[19] I have spent over two years working with the Steering Committee for the AI Index to gather data relating to deployment, and we've found the data to be piecemeal and partial. That's because there are few incentives or mechanisms to get people to describe the systems they deploy, and frequently the main way to know a company or government agency is using an AI system is via a press release from the vendor announcing them as a customer, through media reporting about the product, or through leaks.

while also creating more tools for the federal government to coordinate among itself as it further develops its AI strategy.

3.2 Academia

Academia should carry out more targeted research to deal with problems of process failures, learning algorithm failures, and the union of the two[20]. This will require a combination of directed technical research as well as heavily interdisciplinary research.

The main interventions I think would be useful here are[21]:

- The development of 'playbooks' in partnership with industry and government that can help AI developers avoid process problems.[22]

- Additional funding for interdisciplinary research that brings together multiple academic disciplines to analyze the contexts within which AI algorithms are developed and how these contexts interact with the technical aspects of the systems to cause problems.

- Continued funding for research that seeks to better understand the safety aspects of AI systems and to create tools to more easily interrogate AI systems for traits such as bias, or incorrigible behaviors.

3.3 Industry interventions

Industry, government and academia must engage each other more frequently and comprehensively. While this is a relatively obvious point to make, it bears repeating: I do not think our current conversations are as useful as they could be, nor are they as effective as they could be. My perception of why this is is threefold:

- Government lacks the technical expertise to provide enough touchpoints to industry and academia. By technical expertise, I mean people and institutions tasked with tracking and analyzing technical progress while also gathering data on societal impacts and discussing these findings with industry and academia.

- Academia rarely directly rewards policy engagement by younger students and junior faculty; typically, many tenure-track positions evaluate people for somewhat narrowly scoped work and achievements, and relatively few institutions would heavily weight

[20] We can enable such research via additional funding for academia.
[21] Many of these interventions are currently being carried out by academia, but my observation is that the scale of the issues are sufficiently large we should scale-up funding and activity here significantly.
[22] For an example in another domain, check out the US Digital Service's 'Digital Services Playbook' https://playbook.cio.gov/

policy contributions. (For example, a machine learning professor is predominantly evaluated today on their technical contributions, and typically via participation in the academic publishing system of peer-reviewed papers.)

- Industry tends to be cautious in its interactions with governments, especially when it comes to discussing some of the difficult questions surfaced by AI technology. Such caution makes sense when the government is perceived to lack sufficient personnel to have a detailed discussion, as there are reasonable concerns about misinterpretation leading to adverse policy outcomes.[23]

4. Conclusion

AI is progressing rapidly and, at the same time, it's clear that AI deployment brings both societal and technical challenges. We need to decide as a society what values we apply when developing "good enough" AI and where those values derive from, and we should continue to conduct technical work to give us the tools to better align these systems with societal preferences.

As discussed, I think what we need to address these challenges are:

- More transparency into systems that are being deployed into critical areas of public life.
- Increased government investment to measure, assess, track, and forecast the progress and impacts of AI.
- Greater efforts to make this an interdisciplinary conversation, as the problems are themselves interdisciplinary.

[23] Part of why I am being so blunt here is that the organization I work for does not deploy commercial products into the world, so there is less reason to be cautious during these conversations as we don't have a business that could be impinged on by regulatory actions in response to this testimony.

Jack Clark, Policy Director, OpenAI

Jack Clark is the Policy Director for OpenAI, where he focuses on AI policy and strategy. OpenAI is a San Francisco-based artificial intelligence research company, whose goal is to ensure that increasingly powerful AI systems benefit all of humanity. He is also a Research Fellow at the Center for Security and Emerging Technology (CSET) at Georgetown, and is a member of the Center for a New American Security (CNAS) task force on AI and national security. Jack frequently participates in fact-finding studies and forums relating to AI, including events in recent years with the GAO, the National Academy of Sciences, and the Army Cyber Institute. Jack has given numerous talks about artificial intelligence and its impact on policy, ethics, and security, with recent talks covering AI and dual-use for a CNAS event in 2017, issues of AI policy for the opening keynote of a Princeton conference on AI and Ethics in March 2018, and the social implications of AI progress at the AI for Social Good workshop at the International Conference on Learning Representations (ICLR) in May 2019.

He also helps run the AI Index, an initiative from the Stanford One Hundred Year Study on AI to track and analyze AI progress. In addition, he writes a weekly newsletter about cutting-edge AI research and applications called Import AI (www.importai.net), which is read by read by more than twenty thousand experts around the world.

Chairwoman JOHNSON. Thank you very much. Mx. Joy Buolamwini.

TESTIMONY OF JOY BUOLAMWINI,
FOUNDER, ALGORITHMIC JUSTICE LEAGUE

Mx. BUOLAMWINI. Thank you, Chairwoman Johnson, Ranking Member Baird, and fellow Committee Members, for the opportunity to testify. I'm an algorithmic bias researcher based at MIT. I've conducted studies showing some of the largest recorded racial skin type and gender biases in systems sold by IBM, Microsoft, and Amazon. This research exposes limitations of AI systems that are infiltrating our lives, determining who gets hired or fired, and even who's targeted by the police.

Research continues to remind us that sexism, racism, ableism, and other intersecting forms of discrimination can be amplified by AI. Harms can arise unintended. The interest in self-driving cars is in part motivated by the promise they will reduce the more than 35,000 annual vehicle fatalities. A June 2019 study showed that for the task of pedestrian tracking, children were less likely to be detected than adults. This finding motivates concerns that children could be at higher risk for being hit by self-driving cars. When AI-enabled technologies are presented as lifesavers, we must ask which lives will matter.

In healthcare, researchers are exploring how to apply AI-enabled facial analysis systems to detect pain and monitor disease. An investigation of algorithmic bias for clinical populations showed these AI systems demonstrated poor performance on older adults with dementia. Age and ability should not impede quality of medical treatment, but without care, AI and health can worsen patient outcomes.

Behavior-based discrimination can also occur, as we see with the use of AI to analyze social media content. The U.S. Government is monitoring social media activities to inform immigration decisions despite a Brennan Center report and a USCIS (U.S. Citizenship and Immigration Services) study detailing how such methods are largely ineffective for determining threats to public safety or national security. Immigrants and people in low-income families are especially at risk for having to expose their most sensitive information, as is in the case when AI systems are used to determine access to government services.

Broadly speaking, AI harms can be traced first to privileged ignorance. The majority of researchers, practitioners, and educators in the field are shielded from the harms of AI, leading to undervaluation, de-prioritization, and ignorance of problems, along with decontextualized solutions.

Second, negligent industry and academic norms, there's an ongoing lack of transparency and nuanced evaluations of the limitations of AI.

And third, and overreliance on biased data that reflects structural inequalities coupled with a belief in techno-solutionism. For example, studies of automated risk assessment tools used in the criminal justice system show continued racial bias in the penal system, which cannot be remedied with technical fixes.

We must do better. At the very least, government-funded research on human-centered AI should require the documentation of both included and excluded demographic groups.

Finally, I urge Congress to ensure funding without conflict of interest is available for AI research in the public interest. After co-authoring a peer-reviewed paper testing gender and skin type bias in an Amazon product which revealed error rates of 0 percent for white men and 31 percent for women of color, I faced corporate hostility as a company Vice President made false statements attempting to discredit my MIT research. AI research that exposes harms which challenge business interests need to be supported and protected.

In addition to addressing the *Computer Fraud and Abuse Act*, which criminalizes certain forms of algorithmic biased research, Congress can issue an AI accountability tax. A revenue tax of just .5 percent on Google, Microsoft, Amazon, Facebook, IBM, and Apple would provide more than $4 billion of funding for AI research in the public interest and support people who are impacted by AI harms.

Public opposition is already mounting against harmful use of AI, as we see with the recent face recognition ban in San Francisco and a proposal for a Massachusetts Statewide moratorium. Moving forward, we must make sure that the future of AI development, research, and education in the United States is truly of the people, by the people, and for all the people, not just the powerful and privileged. Thank you.

Next, I look forward to answering your questions.

[The prepared statement of Mx. Buolamwini follows:]

United States House Committee on Science, Space and Technology

June 26, 2019

Hearing on

Artificial Intelligence: Societal and Ethical Implications

Written Testimony of

Joy Buolamwini
Founder, Algorithmic Justice League

Masters in Media Arts and Sciences, 2017, Massachusetts Institute of Technology
MSc Education (Learning & Technology), 2014, Distinction, University of Oxford
BS Computer Science, 2012, Highest Honors, Georgia Institute of Technology

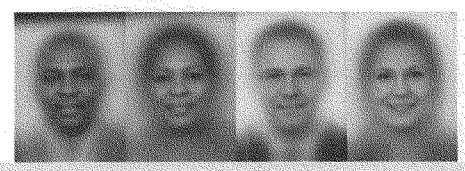
PhD *Pending*, MIT Media Lab

Made Possible By Critical Input from
Dr. Sasha Costanza-Chock
Injoluwa Deborah Raji

For additional information, please contact Joy Buolamwini at joy@ajlunited.org

Dear Chairwoman Johnson, Ranking Member Lucas, and Members of the Committee,

Thank you for the opportunity to testify on the societal and ethical implications of artificial intelligence (AI). My name is Joy Buolamwini, and I am the founder of the Algorithmic Justice League (AJL), based in Cambridge, Massachusetts. I established AJL to create a world with more ethical and inclusive technology after experiencing facial analysis software failing to detect my dark-skinned face until I put on a white mask. I've shared this experience of algorithmic bias in op-eds for Time Magazine and the New York Times as well as a TED featured talk with over 1 million views.[1] My MIT thesis and subsequent research studies uncovered substantial skin type and gender bias in AI services from companies like Microsoft, IBM, and Amazon.[2] This research has been covered in over 40 countries and has been featured in the mainstream media including FOX News, MSNBC, CNN, PBS, Bloomberg, Fortune, BBC, and even the Daily Show with Trevor Noah.[3]



Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Figure 1. Intersectional Skin Type and Gender Classification Accuracy Disparities.
www.gendershades.org

---

[1] The Hidden Dangers of Facial Analysis, New York Times print run June 22, 2018, Page A25, online https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html; Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to Solve It, Time Magazine Optimist Edition http://time.com/5520558/artificial-intelligence-racial-gender-bias/; How I am Fighting Bias in Algorithms, https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms

[2] Joy Buolamwini, Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification (February 2018), http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf; Inioluwa Raji, Joy Buolamwini, Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products (January 2019), http://www.aies-conference.com/wp-content/uploads/2019/01/AIES-19_paper_223.pdf

[3] See references of notable press mentions at www.poetofcode.com/press

Today, I speak to you as both a researcher and someone who has personally experienced algorithmic bias from flaws in AI systems and corporate hostility for publishing research showing gender and racial bias in an existing AI product.

In my testimony today, I will make 5 main points:

- First, the proliferation of AI in society across key social and economic areas makes it nearly impossible for individuals to avoid AI systems, and thus government and academia have an urgent responsibility to address the limitations of AI systems that can mask and further systematize structural inequalities.

- Second, harms from AI systems can arise from systems that propagate error (in)equity such that failures disproportionately impact select groups (i.e pedestrian tracking AI models failing more on children than adults) and processes that create a high exclusion overhead for individuals who fit outside of assumed norms (ie. trans* drivers being forced to undergo continuous and burdensome identification checks and ultimately denied economic opportunity).

- Third, the ability for AI systems to propagate sexism, racism, ableism, and ageism is documented and already marginalized groups like communities of color, low-income families, immigrants and people with disabilities are especially at risk for being further marginalized by AI systems used for employment, healthcare, government services, and policing.

- Fourth, sources of AI harms and bias can arise from lack of diversity in the field, misleading standard benchmarks, data collection and analysis processes, single-axis evaluation norms, and deprioritization of the public interest in the AI development, research, and education.

- Fifth and finally, government and academia must take actions to increase public awareness about the harms of AI, change academic and industry practices that obscure AI limitations, invest in diversifying the field, and ensure research on ethics, accountability, transparency and fairness in AI retains autonomy.

The Proliferation of AI in Society

We have arrived in the age of automation overconfident and underprepared. Often presented as a signifier of progress, artificial intelligence (AI) is increasingly influencing the lives of everyday people in ways that perpetuate individual and societal harms and can amplify past and present-day discrimination. Despite the danger that AI will entrench and exacerbate existing social inequalities, the promise of economic growth coupled with technological advances has spurred widespread adoption. In assessing the economic reach of AI, a recent Mckinsey report states "AI could potentially deliver additional economic output of around $13 trillion by 2030,

boosting global GDP by about 1.2 percent a year."[4] The public sector is also rapidly adopting AI to automate decision making, enhance judgement, improve civic engagement, and streamline interactions with common social services.[5] Taken together, the public and private sector embrace of AI makes it increasingly difficult to function in American society without encountering some form of this technology in consumer products or public services.

Even if an individual attempts to opt-out of an AI-fueled world, their neighbor may install a device with facial recognition enabled surveillance[6], or a bystander may upload a photograph of them to an online platform;[7] they may need to navigate streets increasingly populated with autonomous vehicles[8], submit a resume to an employer using undisclosed and unaccountable automated screening tools,[9] or otherwise interact with automated decision support systems that have already been shown by researchers to be biased and that violate privacy. As I will address more thoroughly already marginalized communities are often further marginalized by the use of these systems.

Select Examples of AI Harms

Though noble intentions like reducing fatalities and overcoming human biases animate the development of AI along with economic interests, research studies and headlines continue to remind us that AI applications are often imbued with bias that can lead to harms.

In identifying AI harms, we must pay particular attention to **error in(equity),** which arises when differential performance across demographics and phenotypic groups leads to harmful bias that disproportionately places the consequences of malfunctions on already marginalized or vulnerable populations (ie. purging of voter registration rolls that rely on automated name matching tools that are biased against non-traditionally European names results in limiting participation in democratic society)

---

[4]Bughin et al. "Notes from the AI Frontier : Modeling the Impact of AI on the World Economy", McKinsey Global Institute, (September 2018),

[5] "Essential Insights: Artificial Intelligence Unleashed", Accenture Federal Services, (2018), https://www.accenture.com/_acnmedia/PDF-86/Accenture-Essential-Insights-POV.pdf#zoom=50

[6] Rich Brown, "Nest says Hello with a new doorbell camera" (September 2017), https://www.cnet.com/news/nest-says-hello-with-a-new-doorbell-camera/

[7] Olivia Solon, "Facial recognition's 'dirty little secret': Millions of online photos scraped without consent" (March 2019), https://www.nbcnews.com/tech/internet/facial-recognition-s-dirty-little-secret-millions-online-photos-scraped-n981921

[8] Kirsten Korosec, "Uber reboots its self-driving car program" (December 2018) https://techcrunch.com/2018/12/20/uber-self-driving-car-testing-resumes-pittsburgh/

[9] Dipayan Ghosh, "AI is the future of hiring, but it's far from immune to bias" (October 2017) https://qz.com/work/1098954/ai-is-the-future-of-hiring-but-it-could-introduce-bias-if-were-not-careful/

We need to also attend to the **exclusion overhead** or the experiential differences that can emerge when technology forces certain demographic groups to expend more time, energy, and resources in an attempt to fit into systems that were optimised for a narrow group but used in a universal manner (ie. changing pitch of voice or speaking patterns to use voice recognition system).

INDIVIDUAL HARMS		COLLECTIVE SOCIAL HARMS
ILLEGAL DISCRIMINATION	UNFAIR PRACTICES	
HIRING		LOSS OF OPPORTUNITY
EMPLOYMENT		
INSURANCE & SOCIAL BENEFITS		
HOUSING		
EDUCATION		
CREDIT		ECONOMIC LOSS
DIFFERENTIAL PRICES OF GOODS		
LOSS OF LIBERTY		SOCIAL STIGMATIZATION
INCREASED SURVEILLANCE		
STEREOTYPE REINFORCEMENT		
DIGNITARY HARMS		

Table 1. Potential Harms from Automated Decision Making[10]

ERROR (IN)EQUITY

Transporting Risks: Which Lives Are We Saving?

According to the National Highway Traffic Safety Administration, vehicle fatalities killed an estimated 36,750 people last year in the United States,[11] and there is growing interest in the potential of autonomous vehicles to reduce deaths and increase transportation efficiency. Yet as Meredith Broussard reminds us in her book *Artificial Unintelligence,* the aspirational vision of what AI could potentially be is not an adequate substitute for reality.

Although autonomous vehicles have captured public and investor imagination, and companies like Tesla and Waymo are pushing the technology forward, development is in nascent stages Missteps including sensor driven fatalities, flawed system designs that enable external hijacking, and research showing pedestrian tracking can be less accurate in detecting dark-skinned

---

[10] See full chart: Lauren Smith, "Unfairness By Algorithm: Distilling the Harms of Automated Decision-Making" (December 2017) https://fpf.org/2017/12/11/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/

[11] "Early Estimate of Motor Vehicle Traffic Fatalities in 2018", US Department of Transportation (2018) https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812749

individuals, demonstrate the need for rigorous evaluations of AI-based vehicles that are entering public spaces.

Because autonomous vehicles must interface with humans, understanding the current performance and risks of the human-centered AI systems that inform car navigation (pedestrian tracking), safety features (drowsy driver alert) or passenger interactions (voice commands, biometric authentication) is critical to developing robust evaluation procedures. Furthermore, growing evidence, including the findings from my research on facial analysis systems, shows that human-centered AI products do not work equally well on different human populations. Differential performance across demographics and phenotypic groups can lead to harmful bias that disproportionately places the consequences of malfunctions on already marginalized or vulnerable populations.

At the recent workshop FATE at CVPR, a leading computer vision conference, an Oxford University researcher shared a study where they evaluated the accuracy of pedestrian detection algorithms. They found a statistically significant difference in the miss rate between adults and children across the top 24 performing algorithmic approaches.These findings along with the recent Georgia Tech study[12] shows that skin type influences the accuracy of state-of-the-art pedestrian tracking methods. These findings motivate concerns that autonomous vehicles that are positioned as lifesavers may in fact do the opposite. The Georgia Tech researchers attributed the difference in accuracy to the lack of representation of darker skinned individuals in training datasets used for pedestrian tracking. Training datasets for pedestrian tracking are not unique in having severe demographic skews. In addition, people with disabilities are seldom included in datasets for human-centered AI systems which further propagates ableism.

Thankfully, we are in the early days of AI development, and there is still time to course correct and exercise caution. Without robust evaluation methods to assess algorithmic vulnerabilities with AVs and high standards to evaluate the distribution of harms, keeping unproven technologies parked will preserve lives. When AI enabled technologies are presented as lifesavers, we must ask which lives will be saved? Which lives will matter?

EXCLUSION OVERHEAD

Hiring and Firing Bias: Who Looks the Part? Who Bears the Exclusion Overhead?
Unlike harmful practices explicitly linked to individual biases or systemic discrimination, AI systems are often perceived as being neutral,[13] making it even more challenging to identify and counteract machine-enhanced racism, sexism, ableism, and other harmful intersecting forms of discrimination. AI enabled tools are increasingly marketed as reducing human bias or being bias free. On the surface, this aim is laudable, but we must again separate potential from reality. The emerging use of AI to inform employment decisions demonstrates that even when AI builders

[12] Wilson et al. "Predictive Inequity in Object Detection" (2019) https://arxiv.org/abs/1902.11097
[13] Nicholas Carr , "The Glass Cage: Automation and Us" (2014)https://dl.acm.org/citation.cfm?id=2666139

hope to overcome human bias they may in fact mask the bias under the guise of machine neutrality.

On December 10, 2018, Upturn released a report detailing the integration of AI tools into human resources from screening to promotion and job termination.[14]Hiring intelligence company HireVue, one of the companies highlighted in the report, explicitly markets its products and services as reducing bias and increasing diversity. HireVue allows employers to interview potential job candidates on camera, by using AI to rate videos of each application according to verbal and nonverbal cues.[15] The system is reportedly trained on the current top performers of a company.[16] However, should exemplar employees be largely homogenous, there is a risk that the data-centric AI system learns to filter out applicants based on features protected by civil rights law (such as race or gender) rather than based on applicants' potential abilities to excel at the job. Amazon learned a similar lesson when an internal AI hiring tool developed to increase efficiency was reported to have harmful gender bias after the system was trained on ten years of hiring data. Unlike HireVue, Amazon's internal tool did not use video input - which introduces new additional risks - but was basing its discrimination on the inclusion of certain gendered descriptions. For instance, if the word "women's" and certain women's colleges appeared in candidates' resumes, the tool ranked them lower.[17]

As I wrote in a New York Times op-ed on June 22, 2018, "Given how susceptible facial analysis technology can be to gender and racial bias, companies using HireVue, if they hope to increase fairness, should check their systems to make sure it is not amplifying the biases that informed previous hiring decisions. It's possible companies using HireVue could someday face lawsuits charging that the program had a negative disparate impact on women and minority applicants, a violation of Title VII of the Civil Rights Act." The hope of overcoming bias cannot be a replacement for rigorous evaluations and external accountability. Beyond having companies implement internal bias mitigation processes, there needs to be external testing and validation to assess the use of AI in employment contexts, as well as regulatory oversight by knowledgeable agencies and consequences for those who violate civil rights law.

AI can serve not only as a gatekeeper for employment, but can also take on the role of terminator. For example, Uber has implemented automated authentication tools to verify that drivers on the platform are who they claim to be. The "Real Time ID Check" tool periodically notifies drivers to take images that are automatically compared to existing driver profile data.

---

[14] Miranda Bogen and Aaron Rieke, "Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias." (2018) https://www.upturn.org/reports/2018/hiring-algorithms.

[15] Corporate Financial Institute, "HireVue Interview Guide: How to prepare for a HireVue interview," accessed on 20 May 2019 https://corporatefinanceinstitute.com/resources/careers/interviews/about-hirevue-interview/

[16] https://www.businessinsider.com/hirevue-ai-powered-job-interview-platform-2017-8

[17] Jeffrey Dastin "Amazon scraps secret AI recruiting tool that showed bias against women" (October 2018) https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

However, the system has limitations. On May 20 2019, Mr Willam Fambroug sued Uber for $227,033 in reparations and punitive damages the company automatically deactivated his account with no means to contest the situation.[18] He states in his legal filing:

> "Uber uses face recognition to verify the correct driver is using the correct driver account. It is universally known, face recognition apps have problems recognizing the "Black" skin color... When asked to verify, .. the app does not recognize my selfie. Uber favors whites who have no problem with the app over blacks who do, as shown by the reasons Uber states for my account deactivation."

This is not an isolated incident nor one that extends only to skin pigmentation. Multiple transgender Uber drivers reported that the feature repeatedly locked them out.[19] Uber reportedly deactivated the accounts of transgender drivers,[20] erroneously denying economic opportunity and highlighting how trans* and gender non-conforming people face additional harms from AI-based tools that are not designed to accommodate a broad range of gender identities and expressions.[21] One former transgender driver with a deactivated account shared that over an 18 month period the Uber system requested over 100 checks for account validations, and that they were suspended from the app for photo inconsistencies as a result. We have to keep in mind not just discriminatory outcomes of AI tools but also the experiences of those using these systems.[22]

These checks require that the driver pull over to take the photo, limiting productivity and time to earn money. I use the term the "exclusion overhead" to capture the experiential differences that can emerge when technology forces certain demographic groups to expend more time, energy, and resources in attempting to fit into systems that were optimised for a narrow group but used in a universal manner. Designers and researchers of AI systems must attend to the exclusion overhead and also keep in mind that AI tools can mask and systematize harmful discrimination.

The use of AI in transportation and employment demonstrate just a handful of ways well intentioned AI tools can propagate harms. Table 1. highlights some additional areas where AI

---

[18] "William Fambrough Vs Uber Technology Inc." May 20 2019 civil suit. Details here: https://drive.google.com/open?id=0B_IVlfmguHPNSFczNVNfWUNzYnNHN21OYVFiZXN2dmdSME9F

[19] https://www.them.us/story/trans-drivers-locked-out-of-uber

[20] Jaden Urbi, "Some transgender drivers are being kicked off Uber's app" in CNBC (August 2018) https://www.cnbc.com/2018/08/08/transgender-uber-driver-suspended-tech-oversight-facial-recognition.html

[21] See more about the harms trans* and gender non-conforming people face from automated decision making systems: Sasha Costanza-Chock, "Design Justice, A.I., and Escape from the Matrix of Domination" in Journal of Design and Science (July 2018), https://jods.mitpress.mit.edu/pub/costanza-chock . For the limitations and harms of binary gender classification see: Os Keyes,"The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition" (2018), https://ironholds.org/resources/papers/agr_paper.pdf

[22] Jaden Urbi, "Some transgender drivers are being kicked off Uber's app" in CNBC (August 2018) https://www.cnbc.com/2018/08/08/transgender-uber-driver-suspended-tech-oversight-facial-recognition.html

systems can limit access to opportunity, render undue economic loss, and perpetuate social stigma. Some areas highlighted in the chart like housing, employment, education, and credit lending have federal protections which make it paramount that we develop AI in a manner that doesn't undercut existing protections and that we educate researchers and practitioners on existing laws. Books like Mireille Hildebrandt's "Law for Computer Scientists and Other Folk" offer a primer to help educate computer scientists on legal matters as the scope of their creations impact society writ large.[23]

Other areas that can lead to collective social harms such as stereotype reinforcement and increased surveillance require an increased awareness of how historic inequalities and controlling narratives shape seemingly objective technologies. In her award-winning book *Dark Matters*, Simone Brown underscores how historic and ongoing oppression - particularly antiblackness - shapes present-day surveillance technologies. And as Shoshana Zuboff emphasizes in her book *Surveillance Capitalism*, the data gathering that fuels large technology companies and recent advancement in AI perpetuate power asymmetries in such a manner where participating in everyday life necessitates submitting to invasive tracking. Both Brown and Zuboff offers insights that can help AI practitioners and researchers better understand how the identification, classification, and measuring of individuals can be used for social control and to deepen entrenched inequalities.

The ExCoded: Further Marginalizing the Already Marginalized

AI Systems Reflect Society

Ultimately, society shapes technology and the shape of American society is one which was built on the genocide and displacement of Indigenous peoples; slavery; the oppression of communities of color, one that did not give women full standing as citizens until the 20th century and still contends with gender discrimination, one that propogated scientific racism, one with a technology industry that is prodimantly led by white men, and one that has allowed corporate interests to influence the policy makers meant to advance the public interest. As such, we have a situation where a small largely homogenous group of people are designing the AI technologies that increasingly touch all of our lives. Without interventions that look at how social and historical factors shape AI development, research, and education, we will increase the technical capabilities of AI systems in ways that continue to worsen inequalities.

For example, AI used to determine hiring decisions has been shown to amplify existing gender discrimination. Law enforcement agencies are rapidly adopting predictive policing and risk assessment technologies that have been shown to reinforce patterns of unjust racial discrimination in the criminal justice system[24]. AI systems also determine the information we see

---

[23] "Law for Computer Scientists" https://lawforcomputerscientists.pubpub.org/

[24] Kristian Lum, William Isaac. "To predict and serve?" (October 2016), https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x ; Rashida Richardson et al. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice" (March 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333423

on social media feeds, and can perpetuate misinformation, amplify hate speech, and unwittingly promote the sexualization of very young children[25] when optimized to prioritize attention-grabbing content.[26] In a world where AI systems influence access to opportunity, freedom, and information, we must attend carefully to the risks they pose and to the distribution of benefits and burdens they produce.

How AI Stigmatizes Cultural Signifiers and Online Behavior

In particular, the burdens of AI fall disproportionately on populations that have been historically excluded from exercising power and obtaining full rights due to patriarchy, white supremacy, and other intersecting forms of oppression. For example, studies of natural language processing (NLP) models that are increasingly used to analyze text for sentiment have revealed how these models often reinforces stereotypes[27],negative associations[28], and misunderstandings of culture.[29] Furthermore, the vast majority of NLP models are trained on what is deemed standard English, making these systems especially ill-equipped to deal with patterns of language such as patois or cultural variations that are not legitimated by state power. Despite these issues, government agencies have explored the use of social media content analysis for extreme vetting to determine who is deemed acceptable and who is deemed a threat[30].

Being labeled suspicious either because your patterns of behavior fit outside what has been defined as normal by an AI system inheriting the power norms of a society, because you belong to a stigmatized group, or because you refuse to submit your activities to algorithmic evaluation can impinge opportunities. In a landmark study on algorithmic bias Dr. Latanya Sweeney, the former chief technologist of the FTC, demonstrated that online searches for names coded as African-American were more likely to bring up search ads associated with arrest records regardless of whether or not the individual actually had a record. In doing due diligence, an employer, landlord, or social worker who searches a stigmatized name may be more likely to dismiss an individual simply because of the risk implied by a negatively associated ad.

Moreover, due diligence is now being automated by AI tools. One company, Predictim, provides a service to conduct background checks on babysitters in part by performing a social media analysis to determine risk ratings for bullying, harassment, being disrespectful and having a bad attitude" Parents are notified whether or not a prospective candidate submits to the search, and

---

[25] Max Fisher and Amanda Taub "On YouTube's Digital Playground, an Open Gate for Pedophiles" (June 2019), https://www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.html

[26]The Spread of True and False News Online: http://science.sciencemag.org/content/359/6380/1146

[27] Bolukbasi et al. "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings" (June 2016), https://arxiv.org/abs/1607.06520

[28] Caliskan et al."Semantics derived automatically from language corpora contain human-like biases" (April 2017) https://science.sciencemag.org/content/356/6334/183.abstract

[29] Su Lin Blodge and Brendan O'Connor, "Racial Disparity in Natural Language Processing:A Case Study of Social Media African-American English" (June 2017),https://arxiv.org/pdf/1707.00061.pdf

[30] Aaron Cantú and George Joseph, "Trump's Border Security May Search Your Social Media by 'Tone'" (August 2017), https://www.thenation.com/article/trumps-border-security-may-search-your-social-media-by-tone/

failure to provide access to personal social media accounts can raise suspicion. Malissa Nielsen, a 24-year-old babysitter who stated she had nothing to hide, submitted her social media information and was surprised to find she was flagged, losing her job in the process. The company does not reveal how these determinations are made, despite the impact they can have over life altering decisions on employment. When AI tools attempt to reduce complex language or behaviour patterns to make unsubstantiated inferences about a person or perpetuate cultural stigma, individuals who belong to communities that have been othered and criminalized will suffer most.

AI Risks for Immigrants, Muslims, and Low-Income Families

Furthermore, those who are in vulnerable situations like refugees seeking asylum or those who face large power asymmetries like immigrants seeking visas, are under increased pressure to subject themselves to algorithmic evaluation or be labeled suspicious for daring to protect their privacy or assert their dignity. For example, recently, the Department of Homeland Security began requiring all visa applicants (15 million people per year) to submit email and social media account information, despite USCIS internal evaluations that show the failure of algorithmic analysis of social media to identify risky actors, and over widespread objections about the potential misuse and harms of automated analysis and classification of immigrants based on social media information. The Brenna Center found Muslims are particularly vulnerable to targeting.[31]

Class dynamics also influence the distribution of burdens from AI systems. In her book *Automating Inequality,* Political Scientist Virginia Eubanks highlights how low-income communities have long been used as guinea pigs to test automated decision making tools. She also illuminates how race and class work together to deepen existing inequalities when top-down tools are introduced into social work. Eubanks chronicles the development of a system implemented in the call screening center for the Allegheny County Office of Children, Youth and Families (CYF) child neglect and abuse hotline to to forecast child abuse and neglect, called the Allegheny Family Screening Tool (AFST)[32]. However, the tool is heavily biased towards predicting children in families with the least resources as being abused, and often overlooking serious cases of neglect in more resourced households. As a result, parents from lower income households get more frequently flagged by the CYF, even despite less evidence of child maltreatment, and are thus at greater risk of losing custody of their children.

As a study on algorithmic risk assessment in child services describes, such cases, although involving a complex analysis of pros and cons, are more likely to become detrimental to low income families[33]:

---

[31] For an in-depth overview of DHS monitoring of social media, see the Brennan Center report at https://www.brennancenter.org/publication/social-media-monitoring; For objections to this kind of monitoring, see http://bit.ly/dhs-social-comments-bu.

[32] Virginia Eubanks,"A Child Abuse Prediction Model Fails Poor Families" (January 2018), https://www.wired.com/story/excerpt-from-automating-inequality/

[33] Chouldechova et al. "A case study of algorithm-assisted decision making in child

> "There is a possibility that some communities— such as those in poverty or from particular racial and ethnic groups—will be disadvantaged by the reliance on government administrative data."

A similar phenomenon of undue burden is observed the allocations of social services. In 2003, a California court case ruled in favour of a welfare system requiring the use of fingerprint identification for aid recipients. The system was implemented as a measure against fraudulent or multiple applications for government aid. However, the prosecuting party also claimed that such measures minimized the impact of the program, deterring immigrants and those experiencing poverty, who were more likely uncomfortable with the practice, from participating and receiving the aid they needed. She also presented the case of these vulnerable groups being justified in their discomfort, as the mandatory fingerprinting, exposing their identifiable biometric data, posed a threat to their overall privacy and safety.[34]

How Automated Proxies Amplify Racism in Price Discrimination and Policing

In reflecting societal patterns, designers of AI models have found the use of zip code to be a powerful variable for making inferences about individuals, because where you live can convey critical information about your place in society including socioeconomic factors like income, education, and employment. The use of zip code can also serve as a proxy that makes decision making seem more neutral by obscuring how geographic locations map to demographics, historic oppression, and ongoing inequalities. The obscuring naturing of zip code coupled with its correlation to demographic factors like race have made it an ideal variable to provide a veneer of objectivity,

In the past, the use of zip code has been intentionally employed to limit material resources and opportunity to the already privileged. The practice of redlining has a long legacy in the United States. Building on patterns of racial segregation, redlining was historically used to keep racial minorities in their place. By categorizing specific zip codes as off limits for receiving loans, raising rates on insurance for minority neighborhoods, and gatekeeping particular neighborhoods to dissuade racial integration, the practice of redlining codified by the passage of the National Housing Act of 1934,[35] was explicitly deployed to preserve opportunities for white communities. Today, AI systems that incorporate geolocation data can learn patterns of exclusion and exploitation. Elevating the visibility of the Tiger Mom Tax, a 2015 study found that test preparation services customers in zip codes with a high density of asian residents were being charged twice the price for services as compared to the average price of these services[36].

---

maltreatment hotline screening decisions" (2018), http://proceedings.mlr.press/v81/chouldechova18a/chouldechova18a.pdf

[34] "Sheyko v. Saenz" civil suit details here: https://cite.case.law/cal-app-4th/112/675

[35] Kevin Fox Gotham, "Racialization and the State: The Housing Act of 1934 and the Creation of the Federal Housing Administration" (2000), https://journals.sagepub.com/doi/10.2307/1389798

[36] Vafa et al. "Price Discrimination in The Princeton Review's Online SAT Tutoring Service"(September 2015), https://techscience.org/a/2015090102

Zip codes are also used in predictive policing applications to indicate areas to patrol for crimes. However, the information that is used is based on past information about areas that have been policed. Given that black and brown neighborhoods are overpoliced and crime in other location is not recorded to the same extent, what might on the surface seems like an objective tool for law enforcement instead reinforces the status quo while using AI to legitimize racialized policing practices.

Colorism, Ageism, and Ablesim in AI for Healthcare

In addition to harms that can arise when AI tools learn or reinforce patterns of discimination, these tools can also lead to bad outcomes when differences between individuals are ignored or erased. Making one group the standard, by which all others must fit can counteract the very benefits designers of AI systems hope to achieve. Studies that highlight breakthroughs in AI for specific domains at times use language that suggests universal progress, when the reality shows a different story. In 2017 Stanford University researched released a study announcing a technical breakthrough in assessing melanoma.[37] The AI systems developed by the researchers matched the accuracy rates comparable to that of dermatologists. However the dataset used for evaluation was overwhelmingly composed of lighter-skinned individuals, even though people with darker skin can get skin cancer.[38] If this model were to be commercialized and used to assess individuals for melanoma, people with skin variations not included in the dataset might have serious problems that could be missed. Already individuals with darker skin are less likely to be diagnosed with melanoma until more advanced stages.[39] Building inclusive AI-enabled diagnostics could help reverse this trend, but only if researchers in the field are intentional.

Age and ability are factors that can influence the technical performance of AI systems built for healthcare. In a February 2019 study, researchers demonstrated the existence of algorithmic bias in state-of-the-art facial expression and landmark recognition methods, which affects the performance of these algorithms for older adults with cognitive impairment.[40] Used as is, the algorithms were less accurate on older adults before being specifically adapted to the population showing that when issues are detected mitigation strategies may be employed. However, they found that even when they attempted to train the algorithms to work better on the faces of older adults with dementia, there were still statistically significant differences between older adults with dementia and those without. The study indicates that not all clinical populations will have the same accuracy even when state-of-the art algorithms are applied to in a wide range of potential health applications including clinical assessment of depression, detection of pain in non-communicative individuals, monitoring progression of motor neuron disease, and alternative interfaces for differently abled persons.

---

[37] Esteva et al. "Dermatologist-level classification of skin cancer with deep neural networks" (February 2017), https://www.nature.com/articles/nature21056

[38] Porcia T. Bradford, "Skin Cancer in Skin of Color" (August 2009), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2757062/

[39] Same as above.

[40] Babak Taati et al., "Algorithmic Bias in Clinical Populations — Evaluating and Improving Facial Analysis Technology in Older Adults With Dementia" (February 2019) https://ieeexplore.ieee.org/document/8643365

As the examples above indicate, we cannot afford to assume AI tools will be bias-free or harmless precisely because these tools, when used in the real world, are part of societal processes that have been shaped by racism, sexism, ableism, and other harmful forms of intersecting discrimination.

Addressing Algorithmic Harms and Bias

Since AI systems are influencing all sectors of society and have documented harms that can increase inequality and facilitate mass surveillance, we must design the processes that shape AI development, research, and education to anticipate, identify, mitigate, and redress harms. Organizations like AI Now and Data & Society have conducted extensive studies that demonstrate the need for interdisciplinary research and policy work that take into account the social and historical contexts that shape the design, development, and governance of AI systems.[41] Tools like algorithmic impact assessments and processes for thinking through legislating AI are crucial components for addressing the harms of AI that extend beyond narrow technical solutions. In addition, AI researchers are starting to contend with issues of ethics, fairness, transparency, and accountability, and there are now a growing number of workshops and conferences in this expanding area of research including the Fairness, Accountability and Transparency (FAT*) conference as well as the AI Ethics and Society (AIES) conference. These combined efforts have helped to spotlight societal sources of AI harms and bias as well as highlight failings in the research and development of AI that have masked problems.

If we are not intentional about designing AI systems with equity in mind, we will replicate existing structural inequalities. With this in mind, below I outline some areas of concern with the state of AI in the United States and their implications for AI harms and bias below. I follow up with my personal experience as an algorithmic bias researcher from a severely underrepresented group (black women) in the domain of AI to provide real-world context to these issues I've encountered firsthand.

- **PRIVILEGED IGNORANCE:** The vast majority of researchers, practitioners, and educators in the field are shielded or removed from the harm that can result in the use of AI systems leading to undervaluation, deprioritization, and ignorance of problems along with decontextualized solutions. The communities most likely to be harmed by AI systems are least likely to be involved in the teaching, design, development, deployment, and governance of AI; even when underrepresented individuals enter previously inaccessible spaces, we face existing practices, norms, and standards that require

[41] AI Now Reports https://ainowinstitute.org/reports.html; Jessie Daniels et al., "Advancing Racial Literacy in Tech" (May 2019) https://datasociety.net/wp-content/uploads/2019/05/Racial_Literacy_Tech_Final_0522.pdf;Kadija Ferryman and Mikaela Pitcan, "Fairness in Precision Medicine" (February 2018) https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_.Precision.Medicine.Feb2018.FINAL-2.26.18.pdf

system-wide not just individual change (ie. Well-meaning organization builds tool to automate screening of children who may be abused or neglected only to make it more likely children who are not at risk but come from low-income families to be targeted)[42]

- **MISLEADING EVALUATIONS:** The field suffers from a false sense of universal progress in part due to misleading evaluation norms and industry wide datasets with significant demographic, phenotypic, and geographic skews. The current push for AI fairness research risks establishing new computational approaches that mask societal problems which cannot be addressed through isolated technical solutions (ie. researchers and practitioners evaluate AI performance based on biased gold standard benchmarks.)
- **PROBLEMATIC DATA FLOWS:** A common response to uncovering severe data imbalances is to collect more data; however, how data is collected, categorized, and distributed presents ethical challenges around consent and privacy along with societal challenges with the politics of classifications where social categories like race and gender become reified into the technical systems that increasingly shape society.
- **SINGLE-AXIS ANALYSIS:** Emerging algorithmic bias research tends to focus on a single-axis of discrimination like race or gender in isolation, missing risks for populations who encounter intersecting forms of discrimination like women of color who contend with both racism and sexism working in combination. Without an intersectional lens our understanding on the scope, spread, and impact of AI harms and bias will be limited. (ie. Government funded human-centered AI research fails to require intersectional analysis echoing issues of government funded health studies in the past not requiring clinical studies data to be disaggregated.)
- **EROSION OF PUBLIC INTEREST:** The risks associated with AI harms and bias threaten trust in government agencies as well as the reputation and product acceptability of influential technology companies who are increasingly funding research and influencing policy discussion around ethics, transparency, accountability, and fairness in AI. Without explicit measures to address conflicts of interest and to protect researchers whose work for the public interest are in tension with private interests, public-private partnerships can lose legitimacy and critical AI harms, research may be silenced, sidelined, and/or underfunded. (ie. Amazon sponsors NSF AI Fairness research despite public company hostility to AI Fairness researchers.)

---

[42] Virginia Eubanks, "Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor" (2018)

PRIVILEGED IGNORANCE

My experiences as one of few black women working on algorithmic bias research has shown me firsthand the importance of having people who are impacted by AI harms and bias working in the field. During the 2015-2016 school year as a masters student at MIT, I had the experience of putting on a white mask in order to have my dark-skinned face consistently detected by face tracking software I incorporated into a coding project. The system I built worked fine on my lighter-skinned colleagues. Like many practitioners, I adopted the practice of using preexisting code made available on the internet in order to integrate the face tracking features. Like many code packages that contain AI models, there was no indication that the system might work better on some groups than others. Without having access to the training data or details about the underlying AI model, I was operating in the dark. This experience of coding in white face motivated the research that became my master's thesis. For this work, I evaluated facial analysis systems from leading tech companies including IBM and Microsoft on the task of guessing the reductive binary gender of a face. All companies performed better on lighter faces than darker faces, and all performed better on male-indentified faces than female-identified faces. When I did an intersectional analysis looking at gender and skin type in combination, I found that error rates were no more than 1% for lighter-skinned men but they soared to over 30% for darker skin women in the worst case. The 2018 research paper published from these findings was widely publicized and covered by national and international media. The public attention lead to private sector action with IBM, Microsoft, and other companies operating in the face space referencing the work in relation to developments on their facial analysis services. The attention also gave me the opportunity to speak to practitioners and researchers inside various companies not just the ones I initially audited. I heard one of three stories:

1. **Decision Makers Deprioritize Issues They Deem Irrelevant or Inconvenient:** In some organizations, junior members of teams reported seeing indications of trouble but senior leadership failed to prioritize issues around algorithmic bias. Among the most troubling case I encountered was from the lead of quality assurance for a company, who expressed regret at not testing accuracy on darker-skinned faces because it would have required more effort than deemed necessary to sell the product.
2. **Homogenous Teams Lack Diverse Perspectives:** In other organizations, despite having access to competitive talent pools, the teams were not aware of the extent of demographic and/or phenotypic bias of their face products and had not explored considering skin type as a variable for analysis
3. **Diverse Leadership Does Not Provide Immunity:** In organizations with people of color in executive roles, I learned that some were aware of bias issues through personal experience and were working to counteract the issues. Still, their products remained biased in part because existing state-of-the-art models and readily available data were biased.

In short, market pressures combined with priorities and constraints of those with the power to create the types of products I scrutinized contributed to the selling of biased AI products. The first two kinds of stories emphasize the importance of having more diverse decision makers, researchers, and practitioners when it comes to identifying, prioritizing, and to some extent empathizing with an issue. The third scenario shows that diversity, while necessary in surfacing issues, is a starting point. Counteracting bias in AI requires not just more inclusive hiring practices.It also requires a multi-pronged approach to shift norms, standards, and incentives, as well as to ensure meaningful external oversight, pressure, regulatory intervention, procurement policies, and significant redress mechanisms for communities that are harmed by biased AI.

MISLEADING EVALUATIONS

In addition to diversifying the practitioners, the practices that shape a field particularly those that have been largely homogenous like AI must also be changed. A pressing question I had as I conducted research which uncovered some of the largest recorded gender and skin-type accuracy disparities in commercial facial analysis was, "Despite my own experiences with technical failures with facial analysis technology, why was I continuing to hear about universal breakthrough about research in the area?" In reviewing key papers on advances in the facial analysis research, I found evaluations of performance generally centered on de facto industry benchmarks. **However, the benchmarks that are used to evaluate the performance of AI systems often have significant representational limitations. Impressive performance on a gold standard can indicate advancement, but if the gold standard only includes data from pale males, we have to ask - improvement for who?**

In 2014, Facebook researchers published the landmark Deep Face paper. Using deep learning techniques, they made a significant leap on the performance of the gold standard facial recognition dataset of the time Labeled faces in the Wild (LFW). They recorded a 97.35% accuracy on LFW significantly exceeding the prior top performance.[43] This was a widely recognized breakthrough and demonstrated the effectiveness of using deep neural networks for computer vision. However, work exploring the demographic composition of LFW, found that the benchmark was 77.5% male and 83.5% White.[44] These overwhelming demographic imbalances persist in core datasets across different domains of AI and limit our understanding of the performance of models on populations that are either severely underrepresented or excluded from benchmarks datasets. The table below provides information about notable imbalances by age, gender, and/or skin type for seven prominent face datasets.

---

[43] Yaniv Taigman et al., "DeepFace: Closing the Gap to Human-Level Performance in Face Verification" (June 2014) https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf

[44] Hu Han, Anil K. Jain, "Age, Gender and Race Estimation from Unconstrained Face Images"(July 2014) http://biometrics.cse.msu.edu/Publications/Face/HanJain_UnconstrainedAgeGenderRaceEstimation_MSUTechReport2014.pdf

	Age Group							Binary Gender[45]		Skin Color / Type	
Dataset	**0-3**	**4-12**	**13-19**	**20-30**	**31-45**	**46-60**	**>60**	**Female**	**Male**	**Darker**	**Lighter**
LFW	1.0%	10.6%	25.4%		29.6%		33.4%	22.5%	77.4%	18.8%	81.2%
IJB-C*	**0.0%**	**0.0%**	**0.5%**	**16.2%**	**35.5%**	**35.1%**	**12.7%**	**37.4%**	**62.7%**	**18.0%**	**82.0%**
Pub fig	1.0%	10.8%	55.5%		21.0%		11.7%	50.8%	49.2%	18.0%	82.0%
CelebA	77.8%					22.1%		58.1%	42.0%	14.2%	85.8%
UTKface	8.8%	6.5%	5.0%	33.6%	22.6%	13.4%	10.1%	47.8%	52.2%	35.6%	64.4%
AgeDB	0.1%	0.52%	2.7%	17.5%	31.8%	24.5%	22.9%	40.6%	59.5%	5.4%	94.6%
IMDB-Face	0.9%	3.5%	33.2%	36.5%	18.8%	5.4%	1.7%	45.0%	55.0%	12.0%	88.0%

Table 2. Age, Binary Gender, and Skin Color/Type Distribution of 7 Prominent Face Datasets
Data reproduced from IBM Research Diversity in Faces Report: https://arxiv.org/pdf/1901.10436.pdf
*IJB-C is a US Government Face Dataset
Produced by the National Institute for Standards and Technology

Moving forward, the field must examine the appropriateness of the metrics and benchmarks by which we measure success and make it common practice to disclose the demographic composition of evaluation benchmark datasets to better assess which populations are either underrepresented or excluded. This basic step for transparency will provide a more realistic view of technical progress.

PROBLEMATIC DATA FLOWS

A common response to uncovering severe data imbalances is to collect more data; however, how data is collected, categorized, and distributed presents ethical challenges around consent, privacy, and compensation along with societal challenges with the politics of classifications. Further, the use of data can provide a veneer of neutrality and objectivity that belie the subjective choices made in selecting and analyzing data.Yes, the rapid adoption of AI in recent years has been made possible by the data surge and increased computation power of the 21st century that now fuels machine learning techniques developed in the 20th century. Machine learning has become the ascendant approach to AI, as the gathering of immense data enables people to use learning algorithms to build models aimed at tasks ranging from classifying faces to identifying disease. **For data-centric technology like machine learning enabled AI, data is destiny. Yet the data that is fueling AI is not neutral. For example, when machines learn from historic practices, they can reinforce past inequalities instead of overcoming them.** Sexist hiring managers or discriminatory recruitment methods are replaced by faceless AI tools that unfairly deny economic opportunity with data-driven precision.

[45] No systematic information is yet available about face based biometric identification system failure rates for gender nonconforming, nonbinary gender, agender, and/or transgender people, specifically.

The flow of data in common AI development pipelines introduces bias at multiple points. Given a generalized overview of a machine learning model development pipeline, there are several areas where bias can be introduced along the way. Understanding how data flows in the practice of building AI models can help with identifying points of intervention, but we must also interrogate how the data that enters a pipeline is obtained.

Current data harvesting processes eschew consent and violate expectations of privacy. Returning to the face space, we see that often convenience sampling is used. Given the availability of online images, researchers and companies scrape the internet for photos generally collected without consent and in violation of expectations of privacy, as Adam Harvey demonstrates in the MegaPixel project. Even when consent is given to store data in one context, scope creep can make it tempting and all too easy for companies storing personal photos to use those images for another purpose.[46]

Classification schema can reify social constructs and limit analysis.
The categorization of data with labels to feed into various AI pipelines often relies on existing classification taxonomies for factors like race, which are socially constructed. While these labels can be useful for conducting disparity audits, they can also risk reifying certain categories and limit analysis. For example, the concept of race, which changes over time, and geography does not denote specific stable physical characteristics and there can be significant intraclass variation. As such using race as a category for evaluating human-centered computer vision tasks can yield poor results compared to use of phenotypic characteristics like skin type, which does not exclusively belong to one socially constructed racial group.[47] Similarly, the common use of binary (Male/Female) gender classification in AI systems may systematically erase the existence of trans*, nonbinary, and gender-nonconforming people, with real-world discriminatory impacts.[48]

The labor and labeling practices used to process data can perpetuate inequality.
Adding to the challenges of choosing classification schema, the application of labels from that schema is often facilitated by employing human annotators who introduce their own individual bias into a labelling process, may be unaware of how their efforts are being utilized (as was the case of worker providing labels to power computer vision applications intended for military

---

[46] James Vincent, "A photo storage app used customers' private snaps to train facial recognition AI" in The Verge (May 2019) https://www.theverge.com/2019/5/10/18564043/photo-storage-app-ever-facial-recognition-secretly-trained
[47] Cynthia M. Cook et al., Demographic Effects in Facial Recognition and Their Dependence on Image Acquisition: An Evaluation of Eleven Commercial Systems (February 2019)
[48] Os Keyes, "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition" (November 2018) https://ironholds.org/resources/papers/agr_paper.pdf; Heath Fogg Davis, "Beyond Trans: Does Gender Matter" (September 2018) https://nyupress.org/9781479855407/

operation), and are poorly paid for their effort[49]. Furthermore human annotators and can introduce additional personal bias making it of particular importance to make sure we properly document data sources, labeling processes, and classification scheme limitations.

As data protections vary in different regions, those with the least protections risk the greatest exploitation.
Beyond providing comprehensive documentation of data collection using approaches like Datasheets for Datasets or Data Nutrition Labels,[50] we must also examine currently accepted research and development practices for mass scale data collection that eschew consent and can violate privacy particularly as legislation like GDPR extends protection to digital information for EU citizens. As data protections vary across jurisdiction we must also be aware of dynamics that can amplify AI and Data Colonialism where individuals with the least protection -in particular, those from the Global South - are the most exploited.[51]

SINGLE-AXIS ANALYSIS

In aiming to address issues of AI harms and bias, an interdisciplinary approach is necessary to provide critical social and historical context as AI is applied in various domains. We need to also make sure that research and evaluation mechanisms like technical standards being developed are adapting to incorporate insights from social sciences. In 1989, legal scholar Kimberlé Crenshaw demonstrated that single-axis antidiscrimination protections by race or by gender were insufficient to protect multiply-burdened groups (in particular, Black women) in the courts. She showed courts repeatedly rejected Black women's discrimination claims when they could only prove that they had been discriminated against specifically as Black women - in other words, their claims were not legally actionable unless they could statistically prove that firms were discriminating either against all women, or against all Black people. Building on Crenshaw's insight, a major focus of my algorithmic bias research has been championing the relevance of intersectional analysis in the domain of human-centered AI systems.

As AI systems are being used for cases like law enforcement, housing, or employment, they must be externally evaluated to assess suitability of use on intended populations should there be legislative approval for deployment. Such evaluations cannot rely on a single aggregate metric for accuracy and must be constructed to disaggregate differences between subpopulations, which can be substantial.

---

[49] Mary L. Gray, Siddharth Suri, "Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass" (2019)
[50] Timnit Gebru et al., "Datasheets for Datasets" (April 2019) https://arxiv.org/pdf/1803.09010.pdf
[51] Amy Hawkins,"Beijing's Big Brother Tech Needs African Faces" in Foreign Policy (July 2018) https://foreignpolicy.com/2018/07/24/beijings-big-brother-tech-needs-african-faces/

GENDER CLASSIFIER	TYPE I	TYPE II	TYPE III	TYPE IV	TYPE V	TYPE VI
Microsoft	1.7%	1.1%	3.3%	0%	23.2%	25.0%
Megvii (Face++)	11.9%	9.7%	8.2%	13.9%	32.4%	46.5%
IBM	5.1%	7.4%	8.2%	8.3%	33.3%	46.8%

Table 3. From 2018 Gender Shades Study: Binary-Gender Classification Error Rates on Women by Fitzpatrick Skin Type

For example, when evaluating error rates for the the facial analysis task of binary-gender classification (which does not account for gender nonconforming people, nonbinary people, agender people, and/or transgender people), our 2018 Gender Shades audit showed women with skin types associated with blackness had error rates as high as 47%. In the same study for men with skin-types perceived as white, error rates were no more than .08% in aggregate. The 47% error rate is of note because binary-gender classification has a 50/50 chance of success based on a random guess.

In our follow up 2019 Algorithmic Justice League Actionable Auditing study, my colleagues and I found that even when target companies improved binary-gender classification performance, publicly attributed to improved training data, they still performed better on lighter-skinned than darker-skinned faces, performed better on male-identified faces than female-identified faces, and performed worst on women of color. Even if accuracy disparities are within a few percentage points, differential performance on millions or hundreds of millions of people will impact many individuals. Therefore, to the extent possible, we need to make sure that we cultivate a practice of employing intersectional analysis in how AI is taught, researcherid, and developed.

EROSION OF PUBLIC INTEREST

In researching and remedying with issues around the ethical and societal implications of AI, the public interest must be prioritized ahead of business interests that are incentivized to maximize profitability over other potential outcomes. The risks associated with AI harms and bias threaten trust in government agencies as well as the reputation and product acceptability of influential technology companies who are increasingly funding research and influencing policy discussion around ethics, transparency, accountability, and fairness in AI. Without explicit measures to address conflicts of interests and to protect researchers whose work for the public interest are in tension with private interests, public-private partnerships can

lose legitimacy and critical AI harms research may be discouraged or underfunded. As Harvard Law Professor Yochai Benkler writes in a Nature op-ed on the recent Amazon-NSF partnership:

> "When the NSF lends Amazon the legitimacy of its process for a $7.6-million programme (0.03% of Amazon's 2018 research and development spending), it undermines the role of public research as a counterweight to industry-funded research...Yes, institutions have erected some safeguards. NSF will award research grants through its normal peer-review process, without Amazon's input, but Amazon retains the contractual, technical and organizational means to promote the projects that suit its goals. "[52]

We cannot forget that companies have no obligation to prioritize the public interest and every incentive to use their power and influence to diminish threats to profitability. Earlier this year, I experienced firsthand corporate backlash after publishing a research study alongside, Deborah Raji, an undergraduate researcher at the time. Our research demonstrated that Amazon Rekognition displayed gender and skin-type bias for the task of gender classification. Amazon Web Services' (AWS) general manager of artificial intelligence, Matthew Wood, and vice president of global public policy Michael Punke attempted to discredit the research with verifiably false claims. One false claim was that I had not made my research methodology available. The methodology used for the study stems from my MIT Master's thesis which was made public in 2017. The 2018 peer-reviewed paper that built on that work also made the methods clear and reproducible. The data used for the study is publicly available for non-commercial use. Researchers in companies such as IBM and Microsoft who could not agree to the terms have instead reproduced comparable data and results using the guidelines written in the original paper. The attacks from Amazon prompted over 70 researchers to write an open letter defending the research and calling for Amazon to stop selling their technology to facial recognition.[53]

Dr. Yoshio Bengio, a recent Turing Prize winner and machine learning pioneer, was one of the authors of this open letter who has been particularly vocal about the need to make sure companies do no usurp AI faculty to the detriment of building the academic capacity of the field. We need not only to preserve academic talent, but also preserve space for critical research within AI. Dr. Bengio observed "The fact that company representatives chose to refute the Raji and Buolamwini paper highlights the importance of a rational and open debate, which will hopefully discourage other companies from using similar tactics, and instead encourage them to

---

[52] Yochai Benkler, "Don't let industry write the rules for AI" (May 2019), https://www.nature.com/articles/d41586-019-01413-1

[53] Ali Alkhatib et al., "On Recent Research Auditing Commercial Facial Analysis Technology" (March 2019) https://medium.com/@bu64dcjrytwitb8/on-recent-research-auditing-commercial-facial-analysis-technology-19148bda1832

improve their products appropriately and engage in a constructive dialogue with scientists who work on these issues."[54]

I can attest that as a researcher who has faced corporate hostility for the work I do, seeing the same corporation that publicly attacked my work showing algorithmic bias in one of their controversial products now sponsor government research in AI Fairness is troubling. Without clear mechanisms that address conflicts of interests (perceived or otherwise), I am less inclined to seek NSF funding for the research that falls under AI Fairness. However, the government should be a major source of funding for research that falls in the realm of technology in the public interest which includes funding for AI Fairness research that is given the space to unabashedly speak truth to power.

Though a plethora of problems exist when considering the depth of the ethical and societal implications of AI, we still have time to institute countervailing mechanisms so that the color of your skin or the inferred contents of your character do not limit access to opportunity under the banner of machine neutrality. Recommendations for government and academia to address the concerns explicated above are briefly outlined below and followed with more broadly focused measures:

PRIVILEGED IGNORANCE: as further outlined below increase awareness of AI harms and adopt proven techniques to diversify the field, such as gathering and sharing demographic data, setting public time-bound diversity and inclusion targets; establishing community review boards that provide real-world perspectives and checks

MISLEADING EVALUATIONS: systematically audit benchmarks for demographic and other relevant categories of bias; establish more diverse and inclusive benchmarks; adopthuman analysis of real-world biased outcomes beyond the mere evaluation of models.

PROBLEMATIC DATA FLOWS: *implement* stronger requirements for consensual data use, to minimize the harms of nonconsensual data use by AI researchers and practitioners;require documentation of data collection and classification processes to increase due diligence

SINGLE-AXIS ANALYSIS: where applicable require intersectional analysis in government funded research, establishing intersectional audit norms, require NIST and other government agencies assessing algorithmic performances to conduct intersectional audits and/or establish a partnership with universities or independent certified agencies to conduct such audits

EROSION OF PUBLIC INTEREST: establish fully autonomous funding for ethics, transparency, accountability and fairness research; procurement processes that require all private vendors of AI services to public agencies to comply with ongoing intersectional bias audits; a requirement

---

[54] Dina Bass, "Amazon Schooled on AI Facial Technology By Turing Award Winner" (April 2019) https://www.bloomberg.com/news/articles/2019-04-03/amazon-schooled-on-ai-facial-technology-by-turing-award-winner

for vendors to submit to community review boards that include members of the most-impacted communities; establish better reporting mechanisms for people to share experiences of harm; decriminalized research in the public interest that is currently penalized by the Computer Fraud and Abuse Act.

Broad Recommendations

INCREASE AWARENESS ABOUT AI HARMS AND BIAS

The public is largely unaware of the ways in which AI shapes their lives, and there are few regulations that require disclosure about the use of the technology. Without awareness about the uses, risks, and limitations of AI, we remain at the mercy of entities that benefit from opaque AI systems, even when they propagate structural inequalities and violate civil rights and liberties. Furthermore practitioners tend to be shielded or removed from the work impacts of AI requiring a shift in how we educate current and future AI developers and researchers. Counteracting the harms of AI and ensuring its benefits are more equitably distributed will require making known existing harms.

Ensure that Computer Science Curriculum from K-12 and post secondary institutions alike addresses issues about the societal and ethical implications of AI and emphasizes that the creators of these systems have an obligation to develop AI in a responsible manner. Examples that are used should be based on real-world occurrences of issues and not theoretical abstractions of hypothetical harms in order to make the issues concrete and stress the need for interdisciplinary knowledge.

Resource public interest technology clinics at degree granting institutions with AI-relevant programs. Such clinics can be modeled on public interest law clinics, so that part of AI education includes a requirement for learning about the real-world consequences of algorithmic harms.

Invest in science communication efforts to make accessible the findings of research studies and results on documenting government testing of AI systems. For example, NIST has been charged with developing AI standards for the United States. The studies produced as an output of these efforts should be presented in a manner where non-domain experts can understand the purpose of the research, the limitations of the methods, and the real-world implications of the results. Researchers receiving government support can be incentivized for making efforts to make their research more accessible. Academic institutions should include course or workshops to help researchers become better communicators of their work.

Promote deeper collaborations between AI researchers and organizations that work most closely with communities that are most harmed by algorithmic inequality. Fund university/community partnerships both to study AI harms on marginalized groups, and also to

do participatory design of AI that is rooted in the needs of marginalized communities. Such collaboration will give a much better contextual understanding of the impact of AI on society, and more importantly enable those who are most impacted by AI harms to be part of the process of counteracting these harms.

Promote creative science initiatives that incorporate the arts and media making to reach broad audiences who otherwise may not encounter research-backed information about existing harms of AI that extend beyond science fiction.

CHANGE RESEARCH & INDUSTRY PRACTICES THAT OBSCURE AI HARMS

Lax research standards plague the field such that critical information about the data used in studies is not collected and/or disclosed,and the homogenous demographic composition of key benchmarks and evaluation norms obscures the potential distribution of harms among different populations. Furthermore, AI products and services are sold with little if any information about potential risks, limitations, and out of context use cases.

Academic institutions and government funding agencies can increase expectations by requiring researchers exploring human-center AI to collect demographic and/or other relevant categorical information as well as document the sourcing, labeling, and interpretation of data collected. Documentation standardization efforts like Datasheets for Datasets [55]and Data Nutrition labels[56] provide starting points for considering what kind of information needs to be collected to inform minimum requirements.

Industry and academics developing AI enabled products or general purpose models should document model performance and provide results to inform stakeholders like organizations considering AI integrations, fellow academics, and the general public. Processes like Model Cards for Model Performance[57] provide a baseline template for considering what kind of information needs to be collected at a minimum so stakeholders can make informed decisions.

INVEST IN DIVERSIFYING THE AI FIELD

The lack of diversity in the field of AI is appalling, particularly considering the wide ranging impact of the output of the technologies that are being developed.

Industry, academia, and government should promote known best practices for addressing diversity gaps. At a minimum transparency in the state of the field is needed which

[55] Gebru et al. "Datasheets for Datasets" (2018), https://arxiv.org/abs/1803.09010
[56] Holland et al. "The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards" (2018), https://arxiv.org/abs/1805.03677
[57] Mitchell et al. "Model Cards for Model Reporting" (2019), https://arxiv.org/abs/1810.03993

will involve gathering demographic data; publishing diversity and inclusion goals; publishing a timeline for reaching those goals; releasing at least annually about progress towards goals.

Support affinity groups that are emerging to address diversity gaps such Black in AI and LatinX in AI.

Provide funding for underrepresented employees, students, and academics to participate in industry events and conferences that may be prohibitively expensive.

CREATE AN AI ACCOUNTABILITY FUND TO SUPPORT CRITICAL RESEARCH AND REDRESS HARMS

Instead of having corporations fund the very research that is meant to keep them accountable, an alternative approach for involving corporations in supporting work in the public interest could be the introduction of an AI Accountability Tax that companies deploying AI systems on a significant portion of the US population must pay. If even Alphabet, Microsoft, Amazon, Facebook, IBM, and Apple tech companies paid .5% of annual revenue, the government raise have $4.4 billion. As a tax, instead of a voluntary contribution through corporate partnership, this money would not be contingent on corporate appetites for engaging with issues of the ethical and societal impacts of AI. Companies like Amazon that have found mechanism to avoid paying corporate taxes, would as a result of their reach and influence still have an obligation to the AI Accountability Fund. Further to support efforts like assess and mitigate AI harms bias, companies architect their tools to enable third-party testing i the public interest that does not incur additional costs for researchers doing this work.

Conclusion

Since AI is being integrated into areas of society including healthcare, education, employment, housing, transportation, and criminal justice that have been shaped by unjust histories and practices, government officials, researchers, and practitioners in the field of AI have an increased responsibility to be especially attuned to the limitations of AI systems that can mask and further systematize structural inequalities regardless of intention. **Algorithmic failures are ultimately human failures that reflect the priorities, values, and limitations of those who hold the power to shape technology.**

We must work to redistribute power in the design, development, deployment, and governance of AI if we hope to realize the potential of this powerful advancement and attend to its perils. We must make sure that the future of AI development, research, and education in the United States is truly of the people, by the people, and for all the people, not just the powerful and privileged.

I look forward to answering your questions,
Joy

Biography

Joy Buolamwini is a graduate researcher at the Massachusetts Institute of Technology who researches algorithmic bias in computer vision systems. She founded the Algorithmic Justice League to create a world with more ethical and inclusive technology. Her TED Featured Talk on algorithmic bias has over 1 million views. Her MIT thesis methodology uncovered large racial and gender bias in AI services from companies like Microsoft, IBM, and Amazon. Her research has been covered in over 40 countries, and as a renowned international speaker she has championed the need for algorithmic justice at the World Economic Forum and the United Nations.

She serves on the Global Tech Panel convened by the vice president of European Commission to advise world leaders and technology executives on ways to reduce the harms of A.I. In late 2018 in partnership with the Georgetown Law Center on Privacy and Technology, Joy launched the Safe Face Pledge, the first agreement of its kind that prohibits the lethal application of facial analysis and recognition technology.

As a creative science communicator, she has written op-eds on the impact of artificial intelligence for publications like TIME Magazine and New York Times. In her quest to tell stories that make daughters of diasporas dream and sons of privilege pause, her spoken word visual audit "AI, Ain't I A Woman?" which shows AI failures on the faces of iconic women like Oprah Winfrey, Michelle Obama, and Serena Williams as well as the Coded Gaze short have been part of exhibitions ranging from the Museum of Fine Arts, Boston to the Barbican Centre, UK. A Rhodes Scholar and Fulbright Fellow, Joy has been named to notable lists including the Bloomberg 50, Tech Review 35 under 35, BBC 100 Women, Forbes Top 50 Women in Tech (youngest), and Forbes 30 under 30. Fortune magazine named her "the conscience of the AI revolution". She holds two masters degrees from Oxford University and MIT; and a bachelor's degree in Computer Science from the Georgia Institute of Technology.

Learn more at www.poetofcode.com

Chairwoman JOHNSON. Thank you very much.
Dr. Georgia Tourassi.

TESTIMONY OF DR. GEORGIA TOURASSI, DIRECTOR, HEALTH DATA SCIENCES INSTITUTE, OAK RIDGE NATIONAL LABORATORY

Dr. TOURASSI. Chairwoman Johnson, Ranking Member Baird, and distinguished Members of the Committee, thank you for the opportunity to appear before you today. My name is Georgia Tourassi. I'm a Distinguished Scientist in the Computing and Computational Sciences Directorate and the Director of the Health Data Sciences Institute of the U.S. Department's Oak Ridge National Laboratory in Oak Ridge, Tennessee. It is an honor to provide this testimony on the role of the Department of Energy and its national laboratories in spearheading responsible use of Federal data assets for AI innovation in healthcare.

The dramatic growth of AI is driven by big data, massive compute power, and novel algorithms. The Oak Ridge National Lab is equipped with exceptional resources in all three areas. Through the Department of Energy's Strategic Partnership Projects program, we are applying these resources to challenges in healthcare.

Data scientists at Oak Ridge have developed AI solutions to modernize the National Cancer Institute's surveillance program. These solutions are being implemented across several cancer registries where they are demonstrating high accuracy and improved efficiency, making near real-time cancer incidents reporting a reality.

In partnership with the Veterans Administration, the Oak Ridge National Lab has brought its global leadership in computing and big data to the task of hosting and analyzing the VA's vast array of healthcare and genomic data. This partnership brings together VA's data assets with DOE's world-class high-performance computing assets and scientific workforce to enable AI innovation and improve the health of our veterans. These are examples that demonstrate what can be achieved through a federally coordinated AI strategy.

But with the great promise of AI comes an even greater responsibility. There are many ethical questions when applying AI in medicine. I will focus on questions related to the ethics of data and the ethics of AI development and deployment.

With respect to the ethics of data, the massive volumes of health data must be carefully protected to preserve privacy even as we extract valuable insights. We need secure digital infrastructure that is sustainable and energy-efficient to accommodate the ever-growing datasets and computational AI needs. We also need to address the sensitive issues about data ownership and data use as the line between research use and commercial use is blurry.

With respect to the ethics of AI development and deployment, we know that AI algorithms are not immune to low-quality data or biased data. The DOE national laboratories, working with other Federal agencies, could provide the secure and capable computing environment for objective benchmarking and quality control of sensitive datasets and AI algorithms against community consensus metrics.

Because one size will not fit all, we need a federally coordinated conversation involving not only the STEM (science, technology, en-

gineering, and mathematics) sciences but also social sciences, economics, law, public policy stakeholders to address the emerging domain-specific complexities of AI use.

Last, we must build an inclusive and diverse AI workforce to deliver solutions that are beneficial to all. The Human Genome Project included a program on the ethical, legal, and social implications of genomic research that had a lasting impact on how the entire community from basic researchers to drug companies to medical workers used and handled genomic data. The program could be a model for a similar effort to realize the hope of AI in transforming health care.

The DOE national laboratories are uniquely equipped to support a national strategy in AI research, development, education, and stakeholder coordination that addresses the security, societal, and ethical challenges of AI in health care, particularly with respect to the Federal data assets.

Thank you again for the opportunity to testify. I welcome your questions on this important topic.

[The prepared statement of Dr. Tourassi follows:]

Statement of Georgia Tourassi
Distinguished Scientist
Director, Health Data Sciences Institute
Oak Ridge National Laboratory

Before the
Committee on Science, Space, and Technology
U.S. House of Representatives
June 26, 2019

Hearing on Artificial Intelligence: Societal and Ethical Implications

Chairwoman Johnson, Ranking Member Lucas, and distinguished members of the Committee: Thank you for the opportunity to appear before you today. My name is Georgia Tourassi. I am a Distinguished Scientist in the Computing and Computational Sciences Directorate and the Director of the Health Data Sciences Institute of the U.S. Department of Energy's Oak Ridge National Laboratory (ORNL) in Oak Ridge, Tennessee. It is an honor to provide this testimony on the transformative role and ethics of Artificial Intelligence (AI) in healthcare and the role of the Department of Energy and its national laboratories in spearheading socially responsible AI science and responsible use of the federal data assets for AI innovation.

INTRODUCTION

As a formally trained physicist, engineer, and applied scientist with three decades of AI experience in biomedical discovery and healthcare delivery, both in academia and a national lab, I have observed the growth, subsequent stagnation, and latest renaissance of AI for health. These experiences have shaped my views on the AI topics you are considering today. In my role as director of the Health Data Sciences Institute and my experience as a computational scientist at ORNL, I have a broader perspective on the potential role of the Department of Energy's national laboratories in supporting an energy-efficient, socially responsible, and secure AI national strategy. I will begin with a brief overview of AI activities at Oak Ridge National Laboratory, exemplifying the impact

of AI for our health. Specifically, I will highlight how the Oak Ridge National Laboratory supports the development of a thriving environment for AI advancement by

(i) providing responsible management of national data assets,
(ii) accelerating innovation, and
(iii) supporting a targeted, ethical, and socially diverse research and workforce development strategy.

With examples drawn from healthcare, I will discuss some of the societal implications of AI and the need for research and development (R&D) investments to ensure ethical use of the data and effective and energy-conscious AI technology. Lastly, I will discuss a vision of how the Department of Energy's national laboratories' AI resources, capabilities, and extensive experience can transform healthcare whilst promoting AI innovation, stakeholder engagement, and security of the nation's most sensitive data assets.

THE IMPACT OF AI FOR OUR HEALTH

The explosive growth of AI is driven by the convergence of big data, massive computational power, and novel algorithms. Together, these three pillars enable AI to accumulate, analyze, and automate the delivery of functional knowledge in many application domains.

The national laboratories are a remarkable asset for the nation. Over the past 75 years, they have consistently provided the science and technology needed to address national problems. As a DOE national laboratory and Federally Funded Research and Development Center, ORNL is equipped with exceptional resources in all three of the pillars of AI to support the Department's mission needs. Researchers at Oak Ridge and other national laboratories have a wealth of experience in building predictive models by using supercomputers to solve complex first-principles physics equations. With the steadily increasing power of today's supercomputers and the massive data sets that are becoming available in a variety of areas, we are now in a position to build the same types of predictive models by training these new models using AI. The leadership-class

supercomputers at the DOE national laboratories are a unique asset for this training step.

Through the Department's Strategic Partnership Projects program, we have been given the opportunity to apply our resources to develop novel and promising AI solutions to several challenges in healthcare, as outlined below.

AI for modernization of national cancer surveillance

Supporting efforts to reduce the cancer burden in the United States, the national Surveillance, Epidemiology, and End Results (SEER) Program manages the collection of curated data from population-based cancer registries on cancer incidence, prevalence, survival, and associated health statistics for the advancement of public health. As part of the National Cancer Institute (NCI), the SEER program is tasked with supporting cancer research to improve the understanding of patient care and outcomes in the "real world" beyond the clinical trial setting.

NCI SEER partnered with DOE national laboratories to leverage their capabilities in high-performance computing and AI, with two goals in mind: to develop a more timely, comprehensive, scalable, and cost-effective cancer surveillance program, and to lay the foundation for an integrative data-driven approach to modeling cancer outcomes at population scale. By leveraging ORNL's exascale computing technologies, computational and AI algorithmic advances achieved through the DOE-NCI partnership provide the scientific framework to evaluate and inform personalized therapies, support the development of prospective diagnostics and treatments, and optimize population health outcomes.

The collaboration has had a direct impact on cancer surveillance over the first three years. Deployed AI algorithms have demonstrated high accuracy over 12 SEER cancer registries and millions of pathology reports received at the registries and processed with increased efficiency. The high accuracy and time efficiency of AI are expected to reduce the workload burden for cancer registrars while allowing them to realign their focus on abstracting additional complex variables (e.g., new cancer biomarkers, cancer recurrence) that are currently not possible using AI. Furthermore, by

leveraging AI for "real time" cancer incidence reporting, we are moving towards real-time eligibility assessment of cancer patients for clinical trials. This is key to NCI's goal of expanding the number and representativeness of patients in clinical trials and other research studies.

This effort to modernize the national cancer surveillance program exemplifies the benefits of a federally coordinated strategy to leverage AI, high-performance computing, and sensitive health data assets for real-world application. In addition, the specific application domain enables ORNL computational scientists to make fundamental contributions to the DOE mission with respect to preparing high-performance computing for the exascale as well as advancing the fundamental computational sciences of AI.

AI for intelligent clinical trials matching

ORNL recently participated in The Opportunity Project (TOP) Health Sprint. This was a 14-week effort sponsored by the US Census Bureau, coordinated by the US Department of Health and Human Services, and led by two Presidential Innovation Fellows. ORNL researchers showcased how AI can refine and advance the process of matching cancer patients to promising clinical trials.

Clinical trials have great potential in advancing the standard of care. However, matching patients with clinical trials remains a challenge, mostly due to the unstructured nature of eligibility criteria as well as the clinical documentation. The approach that we proposed leverages the power of AI technologies such as large-scale knowledge graphs and deep learning to bring together cancer registry data, medical ontologies, and clinical trials data to answer complex questions and provide real-time feedback for patients and clinicians on novel experimental treatments that are available to them.

AI for supporting breast cancer diagnosis

In an effort to reduce errors in the analyses of diagnostic images by health professionals, ORNL scientists have been working to understand and improve the cognitive processes involved in medical image interpretation. Our work has potential to improve health outcomes for the hundreds of thousands of American women affected by breast cancer each year. Our ORNL team found that analyses of radiographic breast

images by expert radiologists were significantly influenced by context bias, or the radiologist's previous diagnostic experiences. So lab reserchers designed an experiment aimed at following the eye movements of radiologists at various skill levels to better understand the context bias involved in their individual interpretations of the images. Using the ORNL supercomputers, our team was able to rapidly train deep learning models necessary to make sense of large datasets of eye-tracking and medical imaging data. These findings will be critical in the future training of medical professionals to reduce errors in the interpretation of diagnostic imaging and will inform the future of human-AI interactions and personalized medical decision support and human augmentation going forward.

AI to improve the health of veterans: MVP-CHAMPION

ORNL is supporting the Veterans Health Administration in the analysis of its massive data resources, including the genomic data gathered through the VA's Million Veterans Program (MVP). The DOE-VA partnership, called the MVP-CHAMPION (Computational Health Analytics for Medical Precision to Improve Outcomes Now), began with a focus on the three health problems that affect the largest number of veterans: prostate cancer, cardiovascular disease, and suicide.

ORNL has brought its global leadership in computing and big data, as well as its demonstrated ability to analyze protected health information on a large scale, to the task of analyzing the VA's data stores. Our work has included the development of advanced algorithms and data mining techniques and of a novel data infrastructure that is consistent, regularly updated, flexible, and easily accessible. These efforts are helping VA and academic researchers tackle the country's most complex health challenges while advancing the state of the art in data analysis and computing, providing benefits to the VA, to DOE, and to the nation's veterans.

AI for Medicare and Medicaid: CMS

A few years ago, ORNL applied its resources to create a Knowledge Discovery Infrastructure (KDI) for the Centers for Medicare and Medicaid Services (CMS). This platform enabled comprehensive and longitudinal analysis of the extraordinarily large

healthcare datasets maintained by CMS, which processes billions of claims each year, and identified patterns indicative of waste, fraud, and abuse. ORNL is now collaborating with the CMS Center for Program Integrity in an effort to develop advanced data analytics methods for detecting fraud.

THE HARD TRUTH OF AI

These examples demonstrate the potential of AI to revolutionize healthcare delivery. While machines won't replace physicians and nurses any time remotely soon, they do have enormous potential to assist health professionals and other decision makers with time-critical decisions for the smart delivery of healthcare. Furthermore, such studies will inform human-computer interactions going forward as we use AI to augment and improve human performance.

Still, the hope of AI must be tempered. AI is often exaggerated with hype and unrealistic expectations of universal benefits. This can lead those outside the AI research communities to believe that AI is the silver bullet that will solve any problem and overcome the known, and unknown, challenges facing us now and in the future. The reality of AI is that it is experiencing a renaissance in terms of capabilities, but still faces the necessary reality check that those capabilities have limitations and there are social implications if we fail to transparently address these limitations. With AI's great promise comes an even greater responsibility. As much positive impact as AI can provide to the nation, we must recognize the pitfalls and possible ill-intended uses of this powerful technology. In this section, I will address separately the ethical and societal implications related to (i) sensitive data management and use, and (ii) AI algorithmic development and practice in the real word, both in the context of healthcare.

The Ethics of Data: Access to large amounts of data is fundamental to AI. Furthermore, access to diverse and seemingly irrelevant data offers the most exciting opportunities in many application domains and certainly in health. For example, the adoption of electronic health records has created great opportunities for biomedical research to understand individual and population-level health outcomes over time. However, liberating and providing access to this data is both a technological and a policy challenge. The different facets of healthcare (clinical services data, pharmacy

data, billing claims, insurance) are often siloed, in part due to policy and the intended use of the collected data. To create a richer picture, medical data cannot be siloed and must be combined with other data points, such as those providing context on a patient's living conditions, a well-known "gap" in the data collection system of the healthcare system that has substantial implications for care delivery. To address this gap, there is a big push in the field to include nontraditional datasets, such as socioeconomic data, socio-demographic statistics, environmental data, and even weather data, in predicting patient health trajectories. Broadly speaking, the biomedical research community is looking for ways to understand and describe the human holistically. At the same time, the sheer volume, variability, and sensitive nature of the personal data being collected require newer, extensive, secure, and sustainable computational infrastructure and algorithms.

Driving sustainable, secure, and energy-efficient infrastructure to handle increasing health data and AI computation: There is a pressing need to consider federal investments in centralizing national data assets in an environment that can provide the compute resources, a secure data management infrastructure, and scalable data analytics capabilities. The Department of Energy national laboratories are uniquely positioned to play that role for the nation, given the Department's long history of serving as the steward of large data infrastructures and of the nation's nuclear security enterprise. With their leading role in high-performance computing and their extensive data science and AI capabilities, the national laboratories could serve as a neutral entity, an honest broker for democratizing AI, while providing meaningful and responsible access to sensitive data assets and compute resources. Such an investment is critical to support the continuum of scientific discovery for effective domain-specific application.

Though AI is currently accelerating research across many fields and industries, an outstanding issue for sensitive domains, such as health and medicine, is how to preserve privacy while computing with shared sensitive medical data to obtain relevant insights. Such data includes personally identifiable information, personal health information, intellectual property, and other proprietary details, where potential leaks would have obvious adverse consequences. Removing personal identifiers and

confidential details is insufficient, as an attacker can still make inferences to recover aspects of the missing data. Inference attacks can also jeopardize AI algorithms over shared data by targeting the shared AI model training process and the trained model itself. Indeed, serious threats are encountered in collective AI endeavors that aggregate data from different sources, since the most vulnerable source establishes the overall security level.

At ORNL, we fully understand the need to overcome these threats as we work to fully utilize the wealth of information in shared data, including data within the ORNL secure data environment, and to apply our high-performance computing resources for AI to make otherwise impossible scientific discoveries and technological breakthroughs in sensitive domains. This is an underdeveloped field of research in which R&D investments are well warranted to develop new solutions so that the community can responsibly and privately share sensitive data for aggregated analysis, including training shared AI models.

Aside from the data security concerns, the energy demands of digital infrastructures and compute- and data-hungry AI algorithms pose energy challenges. AI and other computing activities are projected to use over half of the world's energy by 2040.[1,2] We cannot easily predict how we can balance AI's energy demands with AI's demonstrated ability to guide management of our energy resources more efficiently. DOE's national laboratories are well positioned to lead hardware, software, and algorithmic innovations and deliver AI solutions that consume less energy, a challenge DOE is already working to address as part of the Exascale Computing Initiative. Energy-efficient AI is a key to securing our ability to provide sustainable and affordable solutions with benefits to the environment and our national and economic security.

Driving responsible use of sensitive data: The ethical collection and use of human data is an outstanding challenge for the research community. Although we all recognize the scientific value of human data, the debate over data ownership is ongoing

[1] Greenpeace, "Clicking Clean: Who is winning the race to build a green internet?" Washington, DC, 2017.

[2] Semiconductor Industry Association and Semiconductor Research Corporation, "Rebooting the IT revolution: A call to action," September 2015.

in terms of how best to balance the promise of transparent AI innovation with the risks of unethical data handling, intentional or unintentional privacy breaches, and adversarial data use by hostile or malicious actors.

As scientists, it is our responsibility and as a nation it is our obligation to invest in research and development and related policies to support socially responsible design and use of AI. Data is the new currency in all facets of life and most notably in healthcare, disrupting traditional technology transfer and business models as well as blurring the line between research use and commercial use of patient data. To maintain a strong ethical AI framework, we need to answer this fundamental question: Who owns the intellectual property of data-driven AI algorithms in healthcare? The patient? The medical center collecting the data by providing the healthcare services? Or the AI developer? Clearly, no single entity alone could deliver the breakthrough AI technology. We need a federally coordinated conversation involving all of the STEM sciences, social sciences, economics, law, public policy, and patient advocacy stakeholders. The conversation should also address how to revise and continuously update outdated legislation and regulations to address the emerging complexities of an otherwise exciting technological development.

The Ethics of AI Development and Deployment: On the technical front there are pressing questions on data quality, data bias, AI interpretability, AI vulnerability to adversarial use, and human-AI integration to augment and not inadvertently handicap the human. These topics should be driving our federal research and development priorities and investments. Still, one size does not fit all, and the best technical solutions, regulations, and policies will probably vary across application domains. Therefore, we should recognize that science depends on integration of basic and applied activities. We should invest equally in foundational and applied AI activities, in high-risk disruptive ideas and lower risk incremental endeavors, so that we can reap short-term gains while working towards more sustainable long-term benefits.

Driving responsible development of AI technology: AI developers will need to offer solutions that are not only "on-average" accurate but also offer a measure of trustworthiness at the individual decision level. The latter would require a detailed

explanation of each individual decision AI makes, as well as deeper understanding of the conditions under which the technology is exceptionally successful or alarmingly flawed. To understand these intricate issues, scientists need to dive deeper into the possible sources of errors and uncertainty, including biases in the data collected to train the AI algorithms and implicit biases in the way AI is embedded with the human. Unless these issues are addressed and properly regulated by the appropriate accreditation bodies in healthcare (or any other relevant domain facing similar challenges), we will not see real-world value of AI outside the anecdotal studies.

Driving responsible deployment of AI technology: For example, one area of research that deserves attention is how to most effectively integrate AI technology with the human "in the loop." Although there is increasing fear that AI will eliminate jobs, for the most part AI tools have been shown to be effective for narrowly defined, repetitive tasks. AI will not take the jobs of humans in healthcare, at least not any time soon, but it will change those jobs by improving how humans work, making humans more accurate and more efficient. Ultimately, humans and machines will have to work well together. But this synergy won't happen organically, as past health AI experiences have demonstrated. It is important to train healthcare providers in how to use AI responsibly, how to remain vigilant avoiding mistakes of over-reliance when supported by AI, and how ultimately to be knowledgeable users of the technology.

The DOE national laboratories with the support of other federal agencies could provide a secure environment for objective benchmarking of sensitive datasets and AI algorithms against community consensus metrics to detect, monitor, and possibly correct dataset biases or inconsistent AI technology performance.

Driving the training of an inclusive and diverse AI workforce: With respect to addressing the data bias issue, aside from raising awareness and working towards algorithmic advances, an important first step is to acknowledge that socially responsible AI development and deployment starts with an inclusive and diverse AI workforce. When the community of AI developers both reflects the diversity present in the user community and is embedded in that user community from the start, we will be better positioned to anticipate unintended consequences in the real-world use of AI

technology. Therefore, I advocate a multi-pronged approach throughout the AI lifecycle. First, we need to invest in foundational statistical and data sciences. During the development phase, scientists should promote a rigorous statistical framework to monitor for potential biases in the collected data. During the deployment phase, AI developers should implement rigorous quality control, monitoring AI performance across subgroups to confirm robust performance or identify performance gaps. We should work to communicate to AI users openly and clearly what they should expect across various settings, and we should educate users so that they are informed consumers of the technology. For example, responsible use of AI technology should become part of our mainstream digital education. We cannot anticipate every blind spot and we should not blame AI for learning from implicit biases in the data. Humans do too! It is unavoidable that we will generate AI code faster than we can deliver appropriate policies and regulations. By being inclusive, diverse, rigorous, and vigilant, we can mitigate many of the aforementioned risks. We will be able to democratically answer the ethical questions which will confront us as developers, marketers, users, and regulators.

At ORNL, we are particularly proud of our work with area community colleges and universities to develop the next generation of AI-savvy citizens. Every year we provide high school and college students with internship opportunities to learn AI and use the world's largest computer to address pressing biomedical challenges. In partnership with the University of Tennessee, we offer an interdisciplinary data science and engineering graduate program that is supporting the development of a new generation of highly diverse biomedical scientists with compute and data core competencies. This program could serve as a model for other institutions.

TRANSFORMING HEALTHCARE WITH AI: WHY DOE AND WHY NOW?

Our national healthcare system faces some pressing challenges. The average life expectancy in the United States, 79 years, has increased by more than 30 years over the past century.[3] Much of this progress is due to medical advances with treatment strategies against age-related global killers such as cardiovascular disease, cancer, and

[3] http://www.cdc.gov/nchs/fastats/life-expectancy.htm

stroke.[4] Still, medical research has been less successful at prolonging healthy life (i.e., health span). About 80% of Americans older than 65 live with at least one chronic condition, and 50% live with two.[3] As the aging population in the United States is growing rapidly, the incidence of age-related, costly, chronic conditions such as heart disease, cancer, diabetes, and Alzheimer's disease is reaching epidemic proportions. US healthcare spending already comprises far more of the national GDP than any other sector, including defense, education, energy, and transportation.[5] With annual total costs of age-related diseases expected to skyrocket, exceeding $1.5 trillion in the US by 2030 (considering heart disease,[6] Alzheimer's,[7] and cancer[8] alone), the nation has a pressing need to reduce the economic burden of population aging. Spending cuts and revenue increases cannot solve the increasing federal healthcare costs due to aging.[9] Prolonging our lifespan without prolonging our health span is financially unsustainable for our nation.

AI could offer powerful solutions to these challenges. By leveraging our federal data assets, computing capabilities, and AI, we can develop a strategic roadmap to extend health span and rein in healthcare costs by understanding the broad spectrum of all factors impacting well-being. For the first time, we are at the tipping point to map the human genome (i.e., genomic profile), phenome (i.e., physiologic status), and exposome (i.e., physical and social environment) in real time and across the human lifetime. This is an ambitious endeavor that requires transformative scientific advances in data storage, management, and analytics using massive amounts of heterogeneous health data.

Understanding the human genome-phenome-exposome interplay also demands a multidisciplinary team of biologists, physicists, chemists, engineers, mathematicians, computer scientists, and clinicians as well as the cooperation of key federal, academic, and industrial stakeholders. DOE and its national laboratories have the breadth of

---

[4] http://www.nejm.org/doi/full/10.1056/NEJMra1109345
[5] http://www.usgovernmentspending.com/year_spending_2017USbn_18bs2n#usgs302
[6] http://www.cdcfoundation.org/pr/2015/heart-disease-and-stroke-cost-america-nearly-1-billion-day-medical-costs-lost-productivity
[7] http://www.nejm.org/doi/pdf/10.1056/nejmsa1204629
[8] Yabroff, K. Robin, et al. "Economic burden of cancer in the United States: estimates, projections, and future research." Cancer Epidemiology Biomarkers & Prevention 20.10 (2011): 2006-2014.
[9] https://www.cbo.gov/publication/51580

science and engineering expertise, instrumentation tools and sensor technologies, big data science capabilities, and unique computing resources to tackle this grand challenge. Furthermore, DOE national laboratories are well suited to serve as "honest brokers" of health data to facilitate the complex interactions and cooperation among the various biomedical and healthcare delivery stakeholders.

The Human Genome Project, which was initiated by DOE, was a $3.6B federal investment through its completion in 2003.[10] The human genome-phenome-exposome mapping, a far more ambitious endeavor, will require a multiagency, multinational approach, and the required federal investment could be estimated at $20B in today's dollars.

The national impact of this investment cannot be underestimated. Studies estimate that every federal dollar invested to map the human genome returned between $60 and $140 to the US economy. AI for health could have similar economic impact by generating public and private sector jobs, enabling the development of novel measurement and analytics technologies, and spearheading the health span revolution, a data-driven paradigm shift in human health sciences. Moreover, increasing the average health span by 2 to 3 years could reduce federal healthcare costs by as much as $7.1 trillion over 50 years while improving national well-being and extending workforce participation and productivity.[11]

We can draw another lesson from the Human Genome Project. From its earliest stages, it included a program to address the ethical, legal, and social implications (ELSI) of genomic research. This program, funded at 3 to 5% of the total project budget, became a model for ELSI programs around the world, and the National Human Genome Research Institute continues to fund ELSI research.[12] The program had a lasting impact on how the entire community, from basic researchers to drug companies to medical

[10] Hood, Leroy, and Lee Rowen. "The human genome project: big science transforms biology and medicine." Genome medicine 5.9 (2013): 1.
[11] Goldman, Dana P., et al. "Substantial health and economic returns from delayed aging may warrant a new focus for medical research." Health affairs 32.10 (2013): 1698-1705.
[12] https://plato.stanford.edu/entries/human-genome/

workers, used and handled genetic data. Continuing and expanding this research will help to ensure the responsible use of AI for health.

CONCLUDING REMARKS

The rapid adoption of AI has shown great promise across a wide array of domains such as healthcare, transportation, manufacturing, and cyber security, to name a few. It is becoming an integral part of our daily lives, offering solutions to problems as complex as drug discovery and as trivial as guessing what our next word will be when we text. AI can and will transform the national workforce. It can and will cause social and economic changes. Many of these changes will be positive, but AI can also reinforce and perpetuate harmful societal injustices. One thing is clear. As much as the science of AI is about difficult math and powerful computers, the impact of AI is also about the nation and its citizens. AI is a fundamental human endeavor.

Our nation faces a formidable set of challenges: ensuring our national security in a changing world; increasing the availability of clean, reliable, and affordable energy while protecting the environment; improving human health; and enhancing U.S. competitiveness in the global economy by fostering scientific leadership and encouraging innovation. AI is expected to offer solutions to many of these challenges, but the implications of this disruptive change cannot be underestimated. Close attention to the ethical, legal, and societal implications of AI will be required to ensure that its benefits are shared and its risks are managed and minimized.

A cohesive national plan for AI is an imperative need to secure the nation's economic competitiveness and well-being. At the same time, we as a nation have the unique opportunity to create a well-defined, federally coordinated roadmap to engage in ethical AI development that delivers benefits across all private and public sectors, perhaps most notably in improving human health. The DOE national laboratories are uniquely equipped and positioned not only to make substantial contributions in addressing AI-driven opportunities and challenges, but also to support the execution of a national plan by enabling responsible use of the federal data assets for AI innovation. Although the views I offered in my statement are shaped by my scientific experiences in the healthcare space, they are shared lessons across other application domains.

Thank you again for the opportunity to testify. I welcome your questions on this important topic.

Dr. Georgia Tourassi
Director, Health Data Sciences Institute
Group Leader, Biomedical Science, Engineering and Computing Group
Computing, Science, and Engineering Division
Oak Ridge National Laboratory

Dr. Georgia Tourassi is the founding Director of the Health Data Sciences Institute and Group Leader of Biomedical Sciences, Engineering, and Computing at the Oak Ridge National Laboratory (ORNL). Concurrently, she holds appointments as an Adjunct Professor of Radiology at Duke University and the University of Tennessee Graduate School of Medicine and as a joint UT-ORNL Professor of the Bredesen Center Data Sciences Program and Mechanical, Aerospace, and Biomedical Engineering at the University of Tennessee at Knoxville. Her research interests include artificial intelligence for biomedical applications, medical imaging, biomedical informatics, clinical decision support systems, digital epidemiology, and data-driven biomedical discovery. She served on the FDA Advisory Committee and Review Panel on Computer-Aided Diagnosis Devices, and she has served as appointed Charter member and ad hoc reviewer at several NIH Study sections. Her scholarly work has led to 11 US patents and innovation disclosures and more than 250 peer-reviewed journal articles, conference proceedings articles, editorials, and book chapters. Her research has been featured in numerous high-profile publications such as the MIT Science and Technology Review, Oncology Times and the Economist. Dr. Tourassi has served as Associate Editor of the scientific journals Radiology and Neurocomputing and as a Guest Associate Editor of Medical Physics and IEEE Journal of Biomedical and Health Informatics. She is elected Fellow of the American Institute of Medical and Biological Engineering (AIMBE), the American Association of Medical Physicists (AAPM) and the International Society for Optics and Photonics (SPIE). For her leadership in the Joint Design of Advanced Computing Solutions for Cancer initiative, she received the DOE Secretary's Appreciation Award in 2016. In 2017, Dr. Tourassi received the ORNL Director's Award for Outstanding Individual Accomplishment in Science and Technology and the UT-Battelle Distinguished Researcher Award. Dr. Tourassi holds a B.S. degree in Physics from Aristotle University of Thessaloniki, Greece and a Ph.D. in Biomedical Engineering from Duke University.

Chairwoman JOHNSON. Thank you very much. At this point, we will begin our first round of questions, and the Chair recognizes herself for 5 minutes.

My questions will be to all witnesses. This Committee has led congressional discussions and action on quantum science, engineering, biology, and many other emerging technologies over the years. In thinking about societal implications and governance, how is AI similar to, or different from, other transformational technologies, and how should we be thinking about it differently? We'll start with you, Ms. Whittaker.

Ms. WHITTAKER. Thank you, Chairwoman. I think there are many similarities and differences. In the case of AI, as I mentioned in my opening statement and in my written testimony, what you see is a profoundly corporate set of technologies. These are technologies that, because of the requirement to have massive amounts of computational infrastructure and massive amounts of data, aren't available for anyone with an interest to develop or deploy.

When we talk about AI, we're generally talking about systems that are deployed by the private sector in ways that are calibrated ultimately to maximize revenue and profit. So we need to look carefully at the interests that are driving the production and deployment of AI, and put in place regulations and checks to ensure that those interests don't override the public good.

Chairwoman JOHNSON. Mr. Clark.

Mr. CLARK. It's similar in the sense that it's a big deal in the way that 5G or quantum computers are going to revolutionize chunks of the economy. Maybe the difference is that it's progressing much more rapidly than this technology and it's also being deployed at scale much more rapidly. And I think that the different nature of the pace and scale of deployment means that we need additional attention here relative to the other technologies that you've been discussing.

Mx. BUOLAMWINI. I definitely would want to follow up on scale particularly because even though very few companies tend to dominate the field, the technologies that they deploy can be used by many people around the world. So one example is a company called Megvii that we audited that provides facial analysis capabilities. And more than 100,000 developers use that technology. So you have a case where a technology that is developed by a small group of people can proliferate quite widely and that biases can also compound very quickly.

Chairwoman JOHNSON. Yes.

Dr. TOURASSI. So in the context of the panel I would like to focus on the differences between AI and the technologies that you outlined: Quantum computing and others. AI is not simply about computers or about algorithms. It's about its direct application and use by the humans. So it's fundamentally a human endeavor compared to the other technological advances that you outlined.

Chairwoman JOHNSON. Is it ever too early to start integrating ethical thinking and considerations into all AI research, education, or training, or how can the Federal science agencies incentivize early integration of ethical considerations in research and education at universities or even at K through 12 level?

Ms. WHITTAKER. This is a wonderful question. As I mentioned in my written testimony, I think it is never too early to integrate these concerns, and I think we need to broaden the field of AI research and AI development, as many of my co-panelists have said, to include disciplines beyond the technical. So we need to account for, as we say at AI Now, the full stack supply chain accounting for the context in which AI is going to be used, accounting for the experience of the communities who are going to be classified and whose lives are going to be shaped by the systems, and we need to develop mechanisms to include these at every step of decision-making so that we ensure in complex social contexts where these tools are being used that they're safe and that the people most at risk of harm are protected.

Chairwoman JOHNSON. Thank you.

Mr. CLARK. Very briefly, I think NSF can best integrate ethics into the aspect of grantmaking and also how you can kind of gate for ethics on certain grant applications. And additionally, we should put a huge emphasis on K through 12. I think if you look at the pipeline of people in AI, they drop out earlier than college, and so we should reach them before then.

Mx. BUOLAMWINI. We're already seeing initiatives where even kids as young as 5 and 6 are being taught AI, and there's an opportunity to also teach issues with bias and the need for responsibility. And we're also starting to see competitions that incentivize the creation of responsible AI curriculum. Mozilla Foundation is conducting one of these competitions right now at the undergraduate level.

We also need to look at ways of learning AI that are outside of formal education and look at the different types of online courses that are available for people who might not enter the field in traditional ways and make sure that we're also including ethical and responsible considerations in those areas.

Chairwoman JOHNSON. OK. I'm over my time, but go ahead briefly.

Dr. TOURASSI. As I mentioned in my oral and written testimony, the Human Genome Project represents an excellent example of why and how the ethical, social, and legal implications of AI need to be considered from the beginning, not as an afterthought. Therefore, it should follow both of the scientific realm and having dedicated workforce in that particular space with stakeholders from several different entities to certainly protect and remain vigilant in terms of the scientific advances and the deployment implications of the technology.

Chairwoman JOHNSON. Thank you very much. Mr. Baird.

Mr. BAIRD. Thank you, Madam Chair.

Dr. Tourassi, in this Congress the House Science Committee has introduced H.R. 617 the *Department of Energy Veterans Health Initiative Act*, a bill which I am also a cosponsor. I'm also a Vietnam veteran. And that bill directs the DOE to establish a research program in AI and high-performance computing that's focused on supporting the VA by helping solve big data challenges associated with veterans' health care. In your prepared testimony you highlighted Oak Ridge National Laboratory's work with the joint DOE-VA Million Veterans Program or MVP-CHAMPION (Million Veterans Pro-

gram Computational Health Analytics for Medical Precision to Improve Outcomes Now).

So my question is from your perspective what was the collaboration process like with the VA?

Dr. TOURASSI. From the scientific perspective, it has been a very interesting and fruitful collaboration. Speaking as a scientist who spent a couple of decades in clinical academia before I moved to the Department of Energy, I would say that there is a cultural shift between the two communities. The clinical community will always be focused on translational value and short-term gains when the basic scientific community will be focused on not short-term solutions but disruptive solutions with sustainable value.

In that respect, these are two complementary forces, and I applaud the synergy between basic sciences and applied sciences. It is a relay. Without an important application, we cannot drive meaningfully basic science and vice versa.

Mr. BAIRD. Thank you. And continuing on, what do you feel we can accomplish by managing that large database, and what do you think will help in the——

Dr. TOURASSI. This answer applies not only to the collaboration with the Veterans Administration but in general in the healthcare space. Health care is one of the areas that will be most impacted by artificial intelligence in the 21st century. We have a lot of challenges that do have digital solutions that are compute data-intensive and, by extension, energy security and energy consumption is an issue.

In that respect the collaboration between the DOE national laboratories with the exceptional resources and expertise they have in big data management, secure data management, advanced analytics, and with high-performance computing can certainly spearhead the transformation and enable the development and deployment of tools that will have lasting value in the population.

Mr. BAIRD. So thank you. And continuing on, in your opinion who should be responsible for developing interagency collaboration practices when it comes to data sharing and AI?

Dr. TOURASSI. Again, speaking as a scientist, there are expertise distributed across several different agencies, and all these agencies need to come together to discuss how we need to move forward. I can speak for the national laboratories that they are an outstanding place as federally funded research and development entities to serve as stewards of data assets and of algorithms and to facilitate the benchmarking of datasets and algorithms through the lifecycle of the algorithms, serving as the neutral entities, and while using of course metrics that are appropriate for the particular application domain and driven by the appropriate other Federal agencies.

Mr. BAIRD. So one last question then that deals with your prepared testimony. You described the problems that stem from siloed data in health care. So that relates to what you just mentioned, and you also mentioned the importance of integrating nontraditional datasets, including social and economic data. Briefly, I'm running close on time, so do you got any thoughts on that——

Dr. TOURASSI. You asked two different questions. As I mentioned in my testimony, data is the currency not only for AI, not only in

the biomedical space but across all spaces. And in the biomedical space we need to be very respectful about the patient's privacy. And that has created silos in terms of where the data reside and how we share the data. That in some ways delays scientific innovation.

Mr. BAIRD. Thank you. And I wish I had time to ask the other witnesses questions, but I'm out of time. I yield back, Madam Chair.

Chairwoman JOHNSON. Thank you very much. Mr. Lipinski.

Mr. LIPINSKI. Thank you, Madam Chair. Thank you very much for holding this hearing. I think this is something that we should be spending a whole lot more time on. The impact that AI is having and will have in the future is something we need to examine very closely.

I really want to see AI develop. I understand all the great benefits that can come from it, but there are ethical questions that—tremendous number of things that we have not even had to deal with before.

I have introduced the *Growing Artificial Intelligence Through Research*, or *GrAITR Act* here in the House because I'm concerned about the current state of AI R&D here in the U.S. There's a Senate companion, which was introduced by my colleagues Senators Heinrich, Portman, and Schatz. Now, I want to make sure that we do the technical research but also have to do the research and see what we may need to do here in Congress to let—AI devices are developed consistent with our American values.

I have focused a lot on this Committee because I'm a political scientist. I focus a lot on the importance of social science, and I think it's critically important that social science is not left behind when it comes to being funded because social science has applications to so much technology and certainly in AI.

So I want to ask, when it comes to social science research—and I'll start with Ms. Whittaker—what gaps do you see in terms of the social science research that has been done on AI, and what do you think can and should be done and what should we be doing here in Washington about this?

Ms. WHITTAKER. Thank you. I love this question because I firmly agree that we need a much more broad disciplinary approach to studying AI. To date, most of the research done concerning AI is technical research. Social science or other disciplinary perspectives might be tacked on at the end, but ultimately the study of AI has not traditionally been done through a multi- or interdisciplinary lens.

And it's really important that we do this because the technical component of AI is actually a fairly narrow piece. When you begin to deploy AI in contexts like criminal justice or hiring or education, you are integrating technology in domains with their own histories, legal regimes, and disciplinary expertise. So the fields with domain expertise need to be incorporated at the center of the study of AI, to help us understand the contexts and histories within which AI systems are being applied.

At every step, from earliest development to deployment in a given social context, we need to incorporate a much broader range of perspectives, including the perspectives of the communities

whose lives and opportunities will be shaped by AI decision making.

Mr. LIPINSKI. Mr. Clark?

Mr. CLARK. OpenAI, we recently hired our first social scientist, so that's one. We need obviously many more. And we wrote an essay called, "Why AI Safety Needs Social Scientists." And the observation there is that, along with everything Ms. Whittaker said, we should embed social scientists with technical teams on projects because a lot of AI projects are going to become about values, and technologists are not great at understanding human values but social scientists are and have tools to use and understand them. So my specific pitch is to have federally funded Centers of Excellence where you bring social scientists together with technologists to work on applied things.

Mr. LIPINSKI. Thank you. Anyone else?

Mx. BUOLAMWINI. So I would say in my own experience reading from the social sciences actually enabled me to bring new innovations to computer vision. So in particular my research talks about intersectionality, which was introduced by Kimberlé Crenshaw, a legal scholar who is looking at antidiscrimination law, and showed that if you only did single-access evaluation, let's say you looked at discrimination by race or discrimination by gender, people who were at the intersection were being missed.

And I found that this was the same case for the evaluation of the effectiveness of computer vision AI systems. So, for example, when I did the test of Amazon, when you look at just men or women, if you have a binary, if you look at darker skin or lighter skin, you'll see some discrepancies. But when you do an intersectional analysis, that's where we saw 0 percent error rates for white men versus 31 percent error rates for women of color. And it was that insight from the social sciences to start thinking about looking at intersectionality. And so I would posit that we not only look at social sciences being something that is a help but as something that is integral.

Dr. TOURASSI. As a STEM scientist, I do not speak to the gaps in social sciences, but I know from my own work that for AI technology to be truly impactful, the STEM scientists need to be deeply embedded in the application space to work very closely with the users so that we make sure that we answer the right questions, not the questions that we want to answer as engineers.

And in the biomedical space, we need to be thinking not only about social sciences. We need to be thinking about patient advocacy groups as well.

Chairwoman JOHNSON. Thank you very much. Dr. Babin.

Mr. BABIN. Thank you, Madam Chair. Thank you, witnesses, for being here today.

Mr. Clark and Dr. Tourassi, I have the privilege of representing southeast Texas, which includes the Johnson Space Center. And as the Ranking Member of the Subcommittee on Space and Aeronautics, I've witnessed the diverse ways that NASA has been able to use and develop AI, optimizing research and exploration, and making our systems and technology much more efficient.

Many of the new research missions at NASA have been enhanced by AI in ways that were not previously even possible. As a matter

of fact, AI is a key piece to NASA's next rover mission to Mars, and we could see the first mining of asteroids in the Kuiper belt with the help of AI.

I say all of this to feature the ways that AI is used in the area of data collection and space exploration but to highlight private-public partnerships that have led to several successful uses of AI in this field. Where do you see other private-public partnership opportunities with Federal agencies increasing the efficiency and the security using AI? Dr. Tourassi, if you'll answer first, and then Mr. Clark.

Dr. TOURASSI. So absolutely. The DOE national labs, as federally funded research and development entities, we work very closely with industry in terms of licensing and deploying technology in a responsible way. So this is something that is already hardwired in how we do science and how we translate science.

Mr. BABIN. Thank you very much. Mr. Clark.

Mr. CLARK. My specific suggestion is joint work on robustness, predictability, and broadly, safety, which basically decodes to I have a big image classifier. A person from industry and a person from government both want to know if that's going to be safe and it will serve people effectively, and we should pursue joint projects in this area.

Mr. BABIN. Excellent. Thank you very much. And again, same two, what would it mean for the United States if another country were to gain dominance in AI, and how do we maintain global leadership in this very important study and issue? Yes, ma'am.

Dr. TOURASSI. Absolutely it is imperative for our national security and economic competitiveness that we maintain—we are at the leading edge of the technology and we make responsible R&D investments. In an area that I believe that we can lead the world is that we can actually lead not only with the technological advances but with what we talked about, socially responsible AI. We can lead that dialog, that conversation for the whole world.

Mr. BABIN. Excellent.

Dr. TOURASSI. And that differentiates us from other entities investing in this space.

Mr. BABIN. Yes, thank you. Thank you very much. Mr. Clark.

Mr. CLARK. So I agree, but just to sort of reiterate this, AI lets us encode values into systems that are then scaled against sometimes entire populations, and so along with us needing to work here in the United States on what appropriate values are for these systems, which is its own piece of work, as we've talked about, if we fail here, then the values that our society lives under are partially determined by whichever society wins in AI. And so the values that that society in codes become the values that we experience. So I think the stakes here are societal in nature, and we should not think of this as about a technological challenge but how we as a society want to become better. And the success here will be the ability to articulate values that the rest of the world thinks are the right ones to be embedded, so it's a big challenge.

Mr. BABIN. It is a big challenge. If we do not maintain our primacy in this, then other countries who might be a very repressive with less, you know, lofty values that I assume that's what you're talking about, could put these into effect in a very detrimental way.

So thank you very much. I appreciate it, and I yield back, Madam Chair.

Chairwoman JOHNSON. Thank you very much. Ms. Bonamici.

Ms. BONAMICI. Thank you to the Chair and the Ranking Member, but really thank you to our panelists here.

I first want to note that the panel we have today is not representative of people who work in the tech field, and I think that that is something we need to be aware of because I think it's still probably about 20 percent women, so I just want to point that out.

This is an important conversation, and I'm glad we're having it now. I think you've sent the message that it's not too late, but we really need to raise awareness and figure out if there's policies, if we're talking about the societal part. We have here in this country some of the best scientists, researchers, programmers, engineers, and we've seen some pretty tremendous progress.

But over the years we've talked and spoken in this Committee—and I represent a district in Oregon where we've had lots of conversations about the challenges of integrating AI into our society, what's happening with the workforce in that area, but we really do need to understand better the socioeconomic effects and especially the biases that it can create. And I appreciate that you have brought those to our attention, I mean, particularly for people of color.

And as my colleagues on this Committee know, I serve as the Founder and Co-Chair of the congressional STEAM Caucus to advocate for the integration of arts and design into STEM fields. In *The Innovators*, author Walter Isaacson talked about how the intersection of arts and science is where the digital age creativity is going to occur.

STEAM education recognizes the benefits of both the arts and sciences, and it can also create more inclusive classrooms, especially in the K–12 system. And I wanted to ask Mx. Buolamwini—I hope I said your name——

Mx. BUOLAMWINI. Buolamwini.

Ms. BONAMICI. I appreciate that in your testimony you mentioned the creative science initiatives that are incorporating the arts in outreach to more diverse audiences that may never otherwise encounter information about the challenges of AI. And I wonder if you could talk a little bit about how we in Congress can support partnerships between industry, academia, stakeholders to better increase awareness about the biases that exist because until we have more diversity—you know, it's all about what goes in, that sort of algorithmic accountability I think if you will. And if we don't have diversity going into the process, it's going to affect what's coming out, so——

Mx. BUOLAMWINI. Absolutely. So in addition to being a computer scientist, I'm also a poet. And one of the ways I've been getting the word out is through spoken word poetry. So I just opened an art exhibition in the U.K. in the Barbican that's a part of a 5-year traveling art show which is meant to connect with people who might otherwise not encounter some of the issues that are going on with AI.

Something I would love for Congress to do is to institute a public-wide education campaign. Something I've been thinking about is

a project called Game of Tones, product testing for inclusion. So what you could do——

Ms. BONAMICI. Clever name already.

Mx. BUOLAMWINI. So what you could do is use existing consumer products so maybe it's voice recognition, tone of voice, maybe it's what we're doing with analyzing social media feeds, tone of text, maybe it's something that's to do with computer vision, and use that as a way of showing how the technologies people encounter every day can encode certain sorts of problems, and most importantly, what can be done about it. So it's not just we have these issues, but here are steps forward, here are resources——

Ms. BONAMICI. That's great.

Mx. BUOLAMWINI [continuing]. You can reach out——

Ms. BONAMICI. I serve on the Education Committee as well. I really appreciate that.

Ms. Whittaker, your testimony talks about when these systems fail, they fail in ways that harm those who are already marginalized. And you mentioned that we have to encounter an AI system that was biased against white men as a standalone identity. So increasing diversity of course in the workforce is an important first step, but what checks can we put in place to make sure that historically marginalized communities are part of the decision-making process that is leading up to the deployment of AI?

Ms. WHITTAKER. Absolutely. Well, as we—as I discussed in my written testimony and as AI Now's Rashida Richardson has shown in her research, one thing we need to do is look at the how the data we use to inform AI systems is created, because of course all data is a reflection of the world as it is now, and as it was in the past.

Ms. BONAMICI. Right. Right.

Ms. WHITTAKER [continuing]. And the world of the past has a sadly discriminatory history. So that data runs the risk of imprinting biased histories of the past into the present and the future, and scaling these discriminatory logics across our core social institutions.

Ms. BONAMICI. What efforts are being done at this point in time to do that?

Ms. WHITTAKER. There are some efforts. A paper called Datasheets for Datasets created a framework to provide AI researchers and practitioners with information about the data they were using to create AI systems, including information about the collection and creation processes that shaped a given dataset.

In a law review article titled "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice," AI Now's Director of Policy Research, Rashida Richardson, found that in at least 9 jurisdictions, police departments that were under government oversight or investigation for racially biased or corrupt policing practices were also deploying predictive policing technology.

Ms. BONAMICI. That's very concerning.

Ms. WHITTAKER [continuing]. What this means is that corrupt and racist policing practices are creating the data that is training these predictive systems. With no checks, and no national standards on how that data is collected, validated, and applied.

Ms. BONAMICI. Thank you. And I see I've—my time is expired. I yield back. Thank you, Madam Chair.

Chairwoman JOHNSON. Thank you very much. Mr. Marshall.

Mr. MARSHALL. Thank you, Madam Chair.

My first question for Dr. Tourassi, in your prepared testimony you highlighted that the DOE's partnership with the Cancer Institute Surveillance, Epidemiology, and End Results program, can you explain the data collection process for this program and how the data is kept secure? In what ways have you noted the DOE accounts for artificial intelligence ethics, bias, or reliability at this program? And you also mentioned things like cancer biomarkers that AI are currently unable to predict to produce information on this.

Dr. TOURASSI. The particular partnership with the National Cancer Surveillance program is organized as follows. Cancer is a reportable disease in the U.S. and in other developed countries. Therefore, every single cancer case that is detected in the U.S. is recorded in the local registry. When the partnership was established, the partnership included voluntary participation of cancer registries that wanted to contribute their data to advance R&D.

The data resides in the secure data enclave at the Oak Ridge National Lab where we have the highest regulations and accreditations for holding the data. Access to the data is given responsibly to researchers from the DOE complex that have the proper training to access the data, and that's—that is our test bed for developing AI technology.

The first targets of the science was how we can develop tools that help cancer registries become far more efficient in what they do. It's not about replacing the individual. It's actually helping them do something better and faster. So the first set of tools that are deployed are exactly that, to extract information from pathology reports that the cancer registrars have to report on an annual basis to NCI, and we free time for them to devote to other tasks that are far more challenging for artificial intelligence and—such as the biomarker extraction that you talked about.

Mr. MARSHALL. OK. Thank you so much. I'll address my next question to Mr. Clark but then probably open it up to the rest of the panel after that. How do you incentivize developers to build appropriate safety and security into products when the benefits may not be immediately evident to users?

Mr. CLARK. I think technologists always love competing with each other, and so I'm pretty bullish on the idea of creating benchmarks and challenges which can encourage people to enter systems into this. You can imagine competitions for who's got the least biased system, which actually is something you can imagine commercial companies wanting to participate in. You do need to change the norms of the development community so that individual developers see this as important, and that probably requires earlier education and adding an ethics component to developer education as well.

Mr. MARSHALL. OK. Ms. Whittaker, would you like to respond as well?

Ms. WHITTAKER. Absolutely. I would add to what Mr. Clark's points that it's also important to ensure the companies who build and profit from these systems are held liable for any harms. Com-

panies are developing systems that are having a profound impact on the lives and livelihoods of many members of the public. These companies should be responsible for those impacts, and those with the most power inside these companies should be held most responsible. This is an important point, since most AI developers are not working alone, but are employed within one of these organizations, and the incentives and drivers governing their work are shaped by the incentives of large tech corporations.

Mr. MARSHALL. OK, thanks. Yes, Mx. Buolamwini, sorry I missed the introductions there.

Mx. BUOLAMWINI. Buolamwini. You're fine. And so something else we might consider is something akin to public interest law clinics but are meant for public interest technology so that it's part of your computer science or AI education that you're working with a clinic that's also connected to communities that are actually harmed by some of these processes. So it's part of how you come to learn.

Mr. MARSHALL. OK. Thanks. And, Dr. Tourassi, you get to bat cleanup. Anything you want to add?

Dr. TOURASSI. I don't really have anything to add to this question. I think the other panelists captured it very well.

Mr. MARSHALL. Yes, thank you so much, and I yield back.

Chairwoman JOHNSON. Thank you very much. Ms. Sherrill.

Ms. SHERRILL. Thank you. And thank you to all the panelists for coming today.

This hearing is on the societal and ethical implications of AI, and I'm really interested in the societal dimension when it comes to the impact AI is having on the workforce and how it's increasingly going to shape the future of work. So my first question to the panel is what will the shift in AI mean for jobs across the country? Will the shift to an economy increasingly transformed by AI be evenly distributed across regions, across ethnic groups, across men and women? Will it be evenly distributed throughout our job sectors? And how do you see the percentages of how AI is impacting the workforce changing over the years? Which portion of our workforce will be impacted directly by AI and how will that look for society?

Ms. WHITTAKER. Thank you. Well, I think we're already seeing AI impact the workforce and impact what it means to have a job. We're seeing AI integrated into hiring and recruiting. A company called HireVue now offers video interview services that claim to be able to tell whether somebody is a good candidate based on the way they move their face, their micro-expressions, their tone of voice. Now, how this works across different populations and different skin tones and different genders is unclear because this technology is proprietary, and thus not subject to auditing and public scrutiny.

We are seeing AI integrated into management and worker control. A company called Cogito offers a service to call centers that will monitor the tone of voice and the affect of people on the phone and give them instructions to be more empathetic, or to close the call. It also sends their managers a ranking of how they're doing, and performance assessments can then be based on whatever the machine determines this person is doing well or doing poorly.

We're seeing similar mechanisms in Amazon warehouses where workers' productivity rates are being set by algorithms that are

calibrated to continually extract more and more labor. We've actually seen workers in Michigan walk out of warehouses protesting what they consider inhumane algorithmic management.

Overall, we are already seeing the nature of work reshaped by AI and algorithmic systems, which rely on worker tracking and surveillance and leave no room for workers to contest or even consent to the use of such systems. Ultimately, this increases the power of employers, and significantly weakens the power of workers.

Ms. SHERRILL. And what about—and I'll get—we can go back to you, too, and you can go back to the question if you want, Mx. Buolamwini, but what—to what extent is it going to transfer the ability of people to get jobs and get into the workforce?

Mx. BUOLAMWINI. So one thing I wanted to touch upon is how AI is being used to terminate jobs and something I call the exclusion overhead where people who are not designed for the system have to extend more energy to actually be a part of the system. One example comes from several reports of transgendered drivers being kicked off of Uber accounts because when they used a fraud detection system, which uses facial recognition to see if you are who you say you are, given that they present differently, there were more checks required. So one driver reported that over an 18-month period she actually had to undergo 100 different checks, and then eventually her account was deactivated.

On May 20, another Uber driver actually sued Uber for more than $200,000 after having his account deactivated because he had to lighten his photos so that his face could be seen by these systems, and then there was no kind of recourse, no due process and that he couldn't even reach out to say the reason I lightened my photo, right, was because the system wasn't detecting me.

Ms. SHERRILL. It was failing?

Mx. BUOLAMWINI. Yes.

Ms. SHERRILL. And so also in my district I—and this is to the panel again. I've seen our community colleges and polytechnical schools engaging in conversations with businesses about how they can best train workers to meet the new challenges of the AI workforce and provide them with the skills. Structurally, how does secondary education need to adjust to be able to adapt to the changing needs and the changing challenges that you're outlining? How can we better prepare students to enter into this workforce?

Mr. CLARK. I'll just do a very quick point. We do not have the data to say how AI will affect the economy. We have strong intuitions from everyone who works in AI that it will affect the economy dramatically. And so I'd say before we think about what we need to teach children, we need a real study of how it's impacting things. None of us are able to give you a number on employment——

Ms. SHERRILL. And just because I have 6 seconds, what would you suggest to us to focus on in that study?

Mr. CLARK. I think it would be useful to look at the tractability for in-development technologies to be applied at large scale throughout the economy and to look at the economic impacts of existing things like how we've automated visual analysis and what economic impacts that has had because it's been dramatic but we don't have the data from which to talk about it.

Ms. SHERRILL. Thank you. I yield back. Thank you, Madam Chair.

Chairwoman JOHNSON. Thank you very much. Mr. Gonzalez.

Mr. GONZALEZ. Thank you, Madam Chair, and thank you to our panel for being here today on this very important topic.

Mr. Clark, I want to start my line of questioning with you. It's my belief that the United States needs to lead on machine learning and AI if for no other reason for the sake of standards development, especially when you think about the economic race between ourselves and China. One, I guess, first question, do you share that concern; and then two, if yes, what concerns would you have in a world where China is the one that is sort of leading the AI evolution if you will and dictating standards globally?

Mr. CLARK. Yes, I agree. And to answer your second question, I think if you don't define the standard, then you have less ability to figure out how the standard is going to change your economy and how you can change industry around it, so it just puts you behind the curve. It means that your economic advantage is going to be less, you're going to be less well-oriented in the space, and if you don't invest in the people to go and make those standards, then you're going to have lots of highly qualified reasonable people from China making those. And they'll develop skills, and then we won't get to make them.

Mr. GONZALEZ. Yes, thank you. And then, Dr. Tourassi, another question that I have is around data ownership and data privacy. You know, we talk about the promise of AI a lot, and it is certainly there. I don't know that we talk enough about how to empower individuals with control over their data who are ultimately the ones providing the value by—without even choosing to provide all this data. So in your opinion how should we at the Committee level and as a Congress think about balancing that tradeoff between data privacy and ownership for the individual and the innovation that we know is coming?

Dr. TOURASSI. This is actually an excellent question and fundamental in the healthcare space because, in the end, all the AI algorithmic advances that are happening wouldn't be possible if the patients did not contribute their data and if the healthcare providers did not provide the services that collect the data. So in the end who owns the product?

This is a conversation that requires a societal—as a society to have these pointed conversations about these issues and to bring all the different stakeholders into place. Privacy and ownership mean different things to different people. One size will not fit all. We need to have—to build a framework in place so that we can address these questions per application domain, per situation that arises.

Mr. GONZALEZ. Thank you. And sort of—this one's maybe for everybody, sort of a take on that. Deep fakes is something that we've been hearing a little bit more of lately, and I think the risk here is profound where we get into a world where you literally cannot tell the difference between me calling you on the phone physically or a machine producing my voice. So as we think about that, I guess my question would be, how can the NSF or other Federal agencies ensure that we have the tools available to detect these

deep fakes as they come into our society? We'll start with Ms. Whittaker.

Ms. WHITTAKER. Well, I think this is an area where we need much more research funding and much more development. I would also expand the—this answer to include looking at the environments in which such sensational content might thrive. And so you're looking at engagement-driven algorithmic systems like Facebook, Twitter, and YouTube. And I think addressing the way in which those algorithms surface sensational content is something that needs to go hand-in-hand with detection efforts because, fundamentally, there is an ecology that rests below the surface that is promoting the use of these kind of content.

Mr. GONZALEZ. I completely agree. Thank you.

Mr. CLARK. I agree with these points, and I'd just make one point in addition——

Mr. GONZALEZ. Yes.

Mr. CLARK [continuing]. Which is that we need to know where these technologies are going. We could have had a conversation about deep fakes 2 years ago if you look at the research literature——

Mr. GONZALEZ. Yes.

Mr. CLARK [continuing]. And government should invest to look at the literature today because there will be other challenges similar to deep fakes in our future.

Mx. BUOLAMWINI. We also need to invest in AI literacy where you know that there will be people deploying AI in ways that are meant to be intentionally harmful. So I think making sure people have an awareness that deep fakes can exist and other ways of deception that can arise from AI systems exist as well.

Mr. GONZALEZ. Thank you.

Dr. TOURASSI. So adversarial use of AI technology is a reality.

Mr. GONZALEZ. Yes.

Dr. TOURASSI. It's here. Therefore, the investments in R&D and having an entity that will serve as the neutral entity to steward—to be the steward of the technology and the datasets is a very important piece that we need to consider very carefully and make calculated investments. This is not a one-time solution. Something is clean, ready to go. The vulnerabilities will always exist, so we need to have the processes and the entities in place to mitigate the risks.

And I go back to my philosophy. I believe in what Marie Curie said, "There is nothing to be feared, only to be understood." So let's make the R&D investments to understand. Make the most of the potential and mitigate the risks.

Mr. GONZALEZ. Thank you. I yield back.

Chairwoman JOHNSON. Thank you very much. Mr. McNerney.

Mr. MCNERNEY. Well, thank you. I thank the Chairwoman, and I thank the panelists. The testimony is excellent. I think you all have some recommendations that are good and are going to be helpful in guiding us to move forward, but I want to look at some of those recommendations.

One of your recommendations, Ms. Whittaker, is to require tech companies to waive their secrecy. Now, that sounds great, but in practice it's going to be pretty difficult, especially in light of our competition on the international scene with China and other coun-

tries. How do you envision that happening? How do you envision tech companies opening up their trade secrets without losing the—you know, the competition.

Ms. WHITTAKER. Yes, absolutely. And, as I expand on in my written testimony, this isn't—the vision of this recommendation is not simply that tech companies throw open the door and everything is open to everyone. This is specifically looking at claims of trade secrecy that are preventing accountability. Ultimately, we need public oversight, and overly broad claims to trade secrecy are making that extremely difficult. A nudge from regulators would help here.

We need provisions that waive trade secrecy for independent auditors, for researchers examining issues of bias and fairness and inaccuracy, and for those examining the contexts within which AI systems are being licensed and applied. That last point is important. A lot of the AI that's being deployed in core social domains is created by large tech companies, who license this AI to third parties. Call it an ″AI as a service″ business model. These third parties apply tech company AI in a variety of contexts. But the public rarely knows where and how it's being used, because the contracts between the tech companies and the third parties are usually secret.

Even the fact that there is a contract between, say, Amazon and another entity to license, say, facial recognition is not something that the public who would be profiled by such systems would know. And that makes tracing issues of bias, issues of basic freedoms, issues of misuse extremely hard.

Mr. MCNERNEY. Thank you for that answer. Mr. Clark, I love the way you said that AI encodes the value system of its coders. You cited three recommendations. Do you think those three recommendations you cited will ensure a broader set of values would be incorporated in AI systems?

Mr. CLARK. I described them as necessary but not sufficient. I think that they need to be done along with a larger series of things to incur values. Values is a personal question. It's about how we as a society evaluate what fairness means in a commercial marketplace. And I think that AI is going to highlight all of the ways in which our current systems for sort of determining that need additional work. So I don't have additional suggestions beyond those I make, but is suspect they're out there.

Mr. MCNERNEY. And the idea to have NIST create standards, I mean, that sounds a good idea.

Mr. CLARK. Yes, my general observation is we have a large number of great research efforts being done on bias and issues like it, and if we have a part of government convene those efforts and create testing suites, we can create the source of loose standards that other people can start to test in, and it generates more data for the research community to make recommendations from.

Mr. MCNERNEY. Thank you. Mx. Buolamwini, you recommended a 5 percent AI accountability tax. How did you arrive at that figure, and how do you see that being implemented?

Mx. BUOLAMWINI. So this one was a 0.5 percent tax, and you——

Mr. MCNERNEY. Point 5 percent, thank you.

Mx. BUOLAMWINI [continuing]. And you have the *Algorithmic Accountability Act of 2019* that was sponsored by Representative Yvette Clark. And I think it could be something that is added to

that particular piece of legislation. And so the requirement that they specifically have is this would be for companies that are making over $50 million in revenue or average gross, and then also it would either apply to companies that have or possess over one million consumer devices or reach more than one million consumers. So I could see it being integrated into a larger framework that's already about algorithmic accountability.

Mr. MCNERNEY. Thank you. Ms. Whittaker and Mx. Buolamwini, you both advocated—in fact, all of you did—for a more diverse workforce. I've written legislation to do that. It really doesn't go anywhere around here. What's a realistic way to get that done? How do we diversify the workforce here?

Ms. WHITTAKER. I would hope that lawmakers continue to push legislation that would address diversity in tech, because put frankly, we have a diversity crisis on our hands. It has not gotten better; it has gotten worse in spite of years and years of diversity rhetoric and P.R. We're looking at an industry where——

Mr. MCNERNEY. So you think government is the right tool to make that happen?

Ms. WHITTAKER. I think we need to use as many tools as we have. I think we need to mandate pay equity and transparency. We need to mandate much more thorough protections for people who are the victims of sexual harassment in the workplace. This is a problem that tech has. At Google, for example, more than half of the workforce is made up of contract workers. And this is true across all job types, not just janitors and service workers. You have engineers, designers, project managers, working alongside their full-time colleagues, without the privileges of full employment, and thus without the safety to push back against inequity.

I would add that we also need to look at the practice of hiring increasing numbers of contract workers. These workers are extremely vulnerable to harassment and discrimination. They don't have the protection of full-time employees. And you have seen at Google at this point more than half the workforce is made up of contract workers across all job types, so this isn't just janitorial staff or service workers. This is engineers, designers, team leads that don't have the privileges of full employment and thus don't have the safety to push back against inequity.

Mr. MCNERNEY. I've run out of time, so I can't pursue that. I yield back.

Chairwoman JOHNSON. Thank you very much. Miss González-Colón.

Miss GONZÁLEZ-COLÓN. Thank you, Madam Chair. And yes, I have two questions. Sorry, I was running from another markup. Dr. Tourassi, the University of Puerto Rico, Mayaguez Campus, which is in my district, is an artificial intelligence education and research institute. The facility exposes young students to the field of artificial intelligence. Their core mission is to advance knowledge and provide education in artificial intelligence in theory, methods, system, and applications to human society and to economic prosperity.

My question will be, in your view, how can we engage with institutes of higher education to promote similar initiatives or efforts, keeping in mind generating interest in artificial intelligence in

young students from all areas and how can we be secure that what is produced later on is responsible, ethical, and financially profitable?

Dr. TOURASSI. So, as you mentioned, the earlier we start recruiting workforce, our trainees that reflect the actual workforce with education and the diversity that is needed, that is extremely important. When the AI developers reflect the actual user community, then we know that we have arrived. That cannot be achieved only with academic institutions. This is a societal responsibility for all of us.

I can tell how the national laboratories are working in this space. We are enhancing the academic places and opportunities by offering internship opportunities to students who haven't otherwise—they do not come from research institutions, and this is the first time for them that they can work in a thriving research place. So we need to be thinking more outside the box and how we can all work synergistically and continuously on this.

Miss GONZÁLEZ-COLÓN. Thank you. I want to share with you as well that my office recently had a meeting with a representative of this panel organization, and they were commenting of the challenges they have on approaching American manufacturers, specifically car manufacturers on accessible autonomous vehicles. Several constituents with disabilities rely on them or on similar equipment for maintaining some degree of independence and rehabilitation. My question would be, in your view how can we engage that private-sector—you were just talking a few seconds ago—and the manufacturers that—so we not only ensure that artificial intelligence products are ethical and inclusive but provide opportunities for all sectors of the community, in other words, make this working for everyone? How can we arrange that?

Dr. TOURASSI. If I understood your question, you're asking how we can build more effective bridges?

Miss GONZÁLEZ-COLÓN. In your view, yes, it's kind of the same thing.

Dr. TOURASSI. And again, I can speak to how we are building these bridges as national laboratories working with both academic and research institutions, as well as with private industry creating very thriving hubs for researchers to engage in societally impactful science and develop solutions, end-to-end solutions from R&D all the way to the translation of these products. I see the federally funded R&D entities such as national labs being one form of these bridges.

Miss GONZÁLEZ-COLÓN. How can people with disabilities be counted for when we talk about artificial intelligence?

Dr. TOURASSI. Well, as I said, one size will not fit all. It will come down to the particular application domain, so it is our responsibility as scientists to be mindful of that. And while working, deeply embedded in the application space with the other sciences that will educate us on where the gaps are, that's how we can save ourselves from all the blind spots.

Miss GONZÁLEZ-COLÓN. You said in your testimony—you highlighted the importance of an inclusive and diverse artificial intelligence workforce. For you, what is the greatest challenge in the United States of developing this kind of workforce?

Dr. TOURASSI. As a female STEM scientist and often the token woman for the past three decades in the field, the biggest challenge we have is not actually recruiting a diverse set of trainees but also sustaining them in the workforce. And I passionately believe that we need to change our notion of what is leadership. There are different models of leadership and the more we become comfortable with different styles of leadership. In my own group, in my own team, I make sure that I have a very diverse group of researchers, including people with disabilities, doing phenomenal AI research work. So it comes down to not only developing policies but what is our also individual responsibility as citizens.

Miss GONZÁLEZ-COLÓN. Thank you. And I yield back.

Chairwoman JOHNSON. Thank you very much. Mr. Tonko.

Mr. TONKO. Thank you, Chairwoman Johnson, for holding the hearing, and thank you to our witnesses for joining us.

Artificial intelligence is sparking revolutionary change across industries and fields of study. Its benefits will drive progress in health care, climate change, energy, and more. AI can help us diagnose diseases early by tracking patterns of personal medical history. It can help identify developing weather systems, providing early warning to help communities escape harm.

Across my home State of New York, companies, labs, and universities are conducting innovative research and education in AI, including the AI Now Institute at New York University represented here with us today by Co-Founder Meredith Whittaker. Students at Rensselaer Polytechnic Institute in Troy studying machine logic at the Rensselaer AI and Reasoning Lab—work that could transform our understanding of human-machine communication.

IBM and SUNY Polytechnic Institute have formed a groundbreaking partnership to develop an AI hardware lab in Albany focused on developing computer chips and other AI hardware. That partnership is part of a broader $2 billion commitment by IBM in my home State. This work is more than technical robotics. University of Albany researchers are working on ways to detect AI generated deep fake video alterations to prevent the spread of fake news, an issue that has already impacted some of our colleagues in Congress. These researchers are using metrics such as human blinking rates to weed out deep fake videos from authentic ones.

AI presented great benefits, but it is a double-edged sword. In some studies, AI was able to identify individuals at risk for mental health conditions just by scanning their social media accounts. This can help medical professionals identify and treat those most at risk, but it also raises privacy issues for individuals.

We have also seen evidence of data and technical bias that underrepresents or misrepresents people of color in everything from facial recognition to Instagram filters. As a Committee, I am confident that we will continue to explore both the benefits and risks associated with AI, and I look forward to learning more from our witnesses today.

And my question for all panelists is this: What is an example that illustrates the potential of AI? And what is an example that illustrates the risks? Anyone? Ms. Whittaker.

Ms. WHITTAKER. Yes, I will use the same example for both because I think this gives a sense of the double-edged sword of this

technology. Google's DeepMind research lab applied AI technology to reduce the energy consumption of Google's data centers. And by doing this, they claim to have reduced Google's data center energy bill by 40%. They did this by training AI on data collected from these data centers, and using it to optimizing things like when a cooling fan was turned on, and otherwise much more precisely calibrate energy use to ensure maximum efficiency. So here we have an example of AI being used in ways that can reduce energy consumption, and potentially address climate issues.

But we've also seen recent research that exposes the massive energy cost of creating AI systems, specifically the vast computational infrastructure needed to train AI models. A recent study showed that the amount of carbon produced in the process of training one natural language processing AI model was the same as the amount produced by five cars over their lifetimes. So even if AI, when it's applied, can help with energy consumption, we're not currently accounting for the vast consumption required to produce and maintain AI technologies.

Mr. TONKO. Thank you. Anyone else?

Mr. CLARK. Very, very quickly——

Mr. TONKO. Mr. Clark.

Mr. CLARK. One of the big potentials of AI is in health care and specifically sharing datasets across not just, you know, States and local boundaries but eventually across countries. I think we can create global-class diagnostic systems to save people's lives.

Now, a risk is that all of these things need to be evaluated empirically after we've created them for things like bias, and I think that we lack the tools, funding, and institutions to do that empirical evaluation of developed systems safely.

Mr. TONKO. OK. Mx. Buolamwini?

Mx. BUOLAMWINI. Yes, so I look at computer vision systems where I see both cost for inclusion and cost for exclusion. So when you're using a vision system to, say, detect a pedestrian, you would likely want that to be as accurate as possible as to not hit individuals, but that's also the same kind of technology you could put on a drone with a gun to target an individual as well. So making sure that we're balancing the cost of inclusion and the cost of exclusion and putting in context limitations where you say there are certain categorical uses we are not considering.

Mr. TONKO. Thank you. And Dr. Tourassi, please?

Dr. TOURASSI. Yes. I agree with Mr. Clark that in the healthcare space the promise of AI is evident with clinical decision support systems, for example, for reducing the risk of medical error in the diagnostic interpretation of systems. However, that same field that shows many great examples is full of studies that overhype expectations of universal benefits because these studies are limited to one medical center, to a small population.

So we need to become, as I said, educated consumers of the technology and the hype, the news that are out there. We need to be asking these questions, how extensively this tool has been used, across how many populations, how many States, how many—when we dive into the details and we do that benchmarking that Mr. Clark alluded to, then we know that the promise is real. And there are studies that have done that with the rigor required.

Mr. TONKO. Thank you so much. And with that, I yield back, Madam Chairwoman.

Chairwoman JOHNSON. Thank you very much. Mr. Beyer.

Mr. BEYER. Thank you, Madam Chair. And thank you for holding this hearing. I really want to thank our four panelists for really responsible, credible testimony. I'm going to save all of these printed texts and share them with many friends.

You know, the last 4 years on the Science Committee, AI has come up again and again and again. And we've only had glancing blows at the ethics or the societal implications. We've mostly been talking the math and the machine learning and the promise. Even yesterday, we had Secretary Rick Perry—I can't remember which department he represents, but he was here yesterday—just kidding—raving about artificial intelligence and machine learning.

And thanks, too, for the concrete recommendations; we don't often always get that in the Science Committee. But I counted. There were 24 concrete recommendations that you guys offered, everything from waiving trade secrecy to benchmarking machine learning for its societally harmful failures to even an AI tax, which my friends on Ways and Means will love.

But the one ethical societal failure that we haven't talked about is sort of driven by everything you did. One of your papers talked about the 300,000-time increase in machine-learning power in the last 5 or 6 years compared to Moore's law, which would have been 12 times in the same time. In Virginia, we have something like 35,000 AI jobs we're looking to fill right now. And one of the other papers talked about awareness. And we have certainly had computer scientists here in the last couple of years who talked about ambition awareness.

So let me ask the Skynet question. What do you do about the big picture when—well, as my daughter already says, Wall Street is almost completely run right now by machine learning. I mean, it's all algorithms. I visited the floor of the New York Stock Exchange a couple weeks ago with the Ways and Means Committee, and there were very few people there. The people all disappeared. It's all done algorithmically.

So let's talk about the big picture. Any thoughts on the big-picture societal implication of when AI is running all the rest of our lives?

Mr. CLARK. I think it's pretty clear that AI systems are scaling and they're going to become more capable and at some point we'll allow them to have larger amounts of autonomy. I think the responsible thing to do is to build institutions today that will be robust to really powerful AI systems. And that's why I'm calling for large-scale measurement assessments and benchmarking of existing systems deployed today. And that's because if we do that work today, then as the systems change, we'll have the institutions that we can draw on to assess the growing opportunities and threats of these systems. I really think it's as simple as being able to do weather forecasting for this technical progress, and we lack that infrastructure today.

Mr. BEYER. Mx. Buolamwini, I'm going to mispronounce your name, but you're at MIT, you're right next to Steve Pinker at Harvard. They're doing all this amazing work on the evolution of con-

sciousness and consciousness as an emergent property, one you don't necessarily intend, but there it is. Shouldn't we worry about emergent consciousness in AI, especially as we build capacity?

Mx. BUOLAMWINI. I mean, the worry about conscious AI I think sometimes misses the real-world issues of dumb AI, AIs that are not well-trained, right? So when I go back to an example I did in my opening statement, I talk about a recent study that came out showing pedestrian tracking technologies had a higher miss rate for children, right, as compared to adults. So here we were worried about the AIs becoming sentient, and the ones that are leading to the fatalities are the ones that weren't even well-trained.

Mr. BEYER. Well, I would be grateful—among the 24 thoughtful, excellent suggestions you made—and hopefully, we will follow up on many of them or the ones that are congressionally appropriate—is one more that doesn't deal with the kids that get killed, which is totally good, you know, the issues of ageism, sexism, racism that show up, those are all very, very meaningful, but I think we also need to look long term, which is what good leaders do—about the sentience issue and how we protect, not necessarily make sure it doesn't happen but how we protect. And thank you very much for being part of this.

Madam Chair, I yield back.

Chairwoman JOHNSON. Thank you very much. Mr. Lamb.

Mr. LAMB. Thank you, Madam Chairwoman. A couple of you have hit on some issues about AI as it relates to working people both in the hiring process, you know, discriminating against who they're going to hire and bias embedded in what they're doing, as well as the concerns about AI just displacing people's jobs. But I was wondering if any of you could go into a little more detail on AI in the existing workplace and how it might be used to control working people, to worsen their working conditions. I can envision artificial intelligence applications that could sort of interrupt nascent efforts to organize a workplace or maybe in an organized workplace a union that wants to bargain over the future of AI in that workplace but they're not able to access the sort of data to understand what it even is they're bargaining over.

So I don't know if you can give any examples from present-day where these types of things are already happening or just address what we can do to take on those problems as they evolve because I think they're going to come. Thank you.

Ms. WHITTAKER. Thank you. Yes, I can provide a couple of examples, and I'll start by saying that, as my Co-Founder at AI Now Kate Crawford and the legal scholar Jason Schultz have pointed out, there are basically no protections for worker privacy. AI relies on data, and there are many companies and services currently offering to surveil and collect data on workers. And there are many companies that are now offering the capacity to analyze that data and make determinations based on that analysis. And a lot of the claims based on such analysis have no grounding in science. Things like, "is a worker typing in a way that matches the data-profile of someone likely to quit?" Whether or not typing style can predict attrition has not been tested or confirmed by any evidence, but nonetheless services are being sold to employers that claim to be able to make the connection. And that means that even though they're

pseudoscientific, these claims are powerful. Managers and bosses are acting on such determinations, in ways that are shaping people's lives and livelihoods. And workers have no way to push back and contest such claims. We urgently need stronger worker privacy protections, standards that allow workers to contest the determinations made by such systems, and enforceable standards of scientific validation.

I can provide a couple of examples of where we're seeing worker pushback against this kind of AI. Again, I mentioned the Amazon warehouse workers. We learned recently that Amazon uses a management algorithm in their warehouses. This algorithm tracks worker performance based on data from a sensor that workers are required to wear on their wrist, looking at how well workers are performing in relation to an algorithmically-set performance rate. If a worker misses their rate, the algorithm can issue automatic performance warnings. And if a worker misses their rate too many times—say, they have to go to the bathroom, or deal with a family emergency—the algorithm can automatically terminate them. What becomes clear in examining Amazon's management algorithm, is that these are systems created by those in power, by employers, and designed to extract as much labor as possible out of workers, without giving them any possible recourse.

We have also seen Uber drivers striking, around the time of the Uber IPO. In this case, they were protesting a similar technologically-enabled power imbalance, which manifested in Uber arbitrarily cutting their wages without any warning or explanation. Again, we see such tech being used by employers to increase power asymmetry between workers, and those at the top.

A couple of years ago we saw the massive Virginia teachers strike. What wasn't widely reported was one of the reasons for this strike: the insistence by the school district that teachers wear health tracking devices as a condition of receiving health insurance. These devices collect extremely personal data, which is often processed and analyzed using AI.

You've also seen students protesting AI-enabled education, from Brooklyn, to Kansas, and beyond. Many of these programs were marketed as breakthroughs that would enable personalized learning. What they actually did was force children to sit in front of screens all day, with little social interaction or personal attention from teachers.

In short, we've seen many, many examples of people pushing back against the automation of management, and the unchecked centralized power that AI systems are providing employers, at the expense of workers.

Finally, we've also seen tech workers at these companies organizing around many of these issues. I've been a part of a number of these organizing efforts, which are questioning the process by which such systems are created. Tech workers recognize the dangers of these technologies, and many are saying that they don't want to take part in building unethical systems that will be used to surveil and control. Tech workers know that we have almost no checks or oversight of these technologies, and are extremely concerned that they will be used for exploitation, extraction, and

harm. There is mounting evidence that they are right to be concerned.

Mr. LAMB. Thank you very much. I'm just going to ask one more question. I'm almost out of time. Ms. Tourassi—or, Dr. Tourassi, I'm sorry, I know that Oak Ridge has been a partner with the Veterans Health Administration, MVP-CHAMPION I think it's called, and if you could just talk a little bit about—is that project an example of the way that the VA can be a leader in AI as it relates to medicine, precision medicine? You know, we've got this seven-million veteran patient population, and in a number of IT areas we think of it as a leader that can help advance the field. Are you seeing that or are there more things we could be doing?

Dr. TOURASSI. The particular program you described, it's part of the Strategic Partnerships Program that brings the AI and high-performance computing expertise that exist within the DOE national lab system with the application domain and effectively the data owners as well. So that partnership is what's pushing the field forward in terms of developing technologies that we can deploy in the environment of the VA Administration to improve veterans' health care.

I wouldn't consider the Veterans Administration as spearheading artificial intelligence, but, as I said in my written testimony, talent alone is not enough. You need to have the data, you have to—you need to have the compute resources, and you need to have talent. The two entities coming together, they create that perfect synergy to move the field forward.

Mr. LAMB. Well, thank you for that. And I do believe that labs like yours and the efforts that we make in the VHA system are a way that we can help push back against the bias and discrimination in this field because the government really at its best has tried to be a leader in building a diverse workforce of all kinds and allowing workers at least in the Veterans Administration to organize and be part of this whole discussion, so hopefully we can keep moving that forward.

Madam Chair, I yield back. Thank you.

Chairwoman JOHNSON. Thank you very much. Mr. Sherman.

Mr. SHERMAN. Thank you. I've been in Congress for about 23 years, and in every Committee we focus on diversity, economic disruption, wages, and privacy. And we've dealt with that here today as well.

I want to focus on something else that is more than a decade away, and that is that the most explosive power in the universe is intelligence. Two hundred thousand years ago or so our ancestors said hello to Neanderthal. It did not work out well for Neanderthal. That was the last time a new level of intelligence came to this planet, and it looks like we're going to see something similar again, only we are the Neanderthal.

We have, in effect, two competing teams. We have the computer engineers represented here before us developing new levels of intelligence, and we have the genetic engineers quite capable in the decades to come of inventing a mammal with a brain—hundreds of pounds.

So the issue before us today is whether our successor species will be carbon-based or silicon-based, whether the planet will be inher-

ited by those with artificial intelligence or biologically engineered intelligence.

There are those who say that we don't have to fear any computer because it doesn't have hands. It's in a box; it can't affect our world. Let me assure you that there are many in our species that would give hands to the devil in return for a good stock tip.

The chief difference between the artificial intelligence and the genetically engineered intelligence is survival instinct. With DNA, it's programmed in. You try to kill a bug, it seems to want to survive. It has a survival instinct. And you can call it survival instinct; you could call it ambition. You go to turn off your washing machine or even the biggest computer that you've worked with, you go to unplug it, it doesn't seem to care.

What amount of—what percentage of all the research being done on artificial intelligence is being used to detect and prevent self-awareness and ambition? Does anybody have an answer to that? Otherwise, I'll ask you to answer for the record. Yes, sir.

Mr. CLARK. We have an AI safety team at OpenAI, and a lot of that work is about—if I set an objective for a computer, it will probably solve that objective, but it will sometimes do—solve that objective in a way that is incredibly harmful to people because, as other panelists have said, these algorithms are kind of dumb.

Mr. SHERMAN. Right.

Mr. CLARK. What you can do is you can try and have these systems learn values from people.

Mr. SHERMAN. Learning values is nice. What are you doing to prevent self-awareness and ambition?

Mr. CLARK. The idea is that if we encode the values that people have into these systems and so——

Mr. SHERMAN. I don't want to be replaced by a really nice new form of intelligence. I'm looking for a tool that doesn't seek to affect the world.

I want to move onto another issue, related though. I think you're familiar with the Turing test, which in the 1950s was proposed as the way we would know that computers had reached or exceeded human intelligence, and that is could you have a conversation with a computer and not know you're having a conversation with a computer? In this room in 2003 top experts of then predicted that the Turing test would be met by 2028. Does anybody here have a different view? Is that as good an estimate as any? They said it would be 25 years, and that was back in 2003.

I'm not seeing anybody jump up with a different estimate, so I guess we have that one. You're not quite jumping up, but go ahead.

Ms. WHITTAKER. I don't have an estimate on that. I do question the validity of the Turing test insofar as it relies on us to define what a human is, which is of course a philosophical question that we could debate for hours.

Mr. SHERMAN. Well, I don't know about philosophers, but the law pretty well defines who's a human and who isn't and, of course, if we invent new kinds of sentient beings, the law will have to grow.

I just want to add Mr. Beyer brought this up and was kind of dismissed by the idea that we shouldn't worry about a new level of intelligence since we, as of yet, don't have a computer that can drive a car without hitting a child. I think it's important that if

we're going to have computers drive cars that they not hit children, but that's not a reason to dismiss the fact that between biological engineering and computer engineering, we are the Neanderthal creating our own Cro-Magnon.

I yield back.

Chairwoman JOHNSON. Thank you very much. Ms. Horn.

Ms. HORN. Thank you, Madam Chairwoman. And thank you to the panel for an important and interesting conversation today.

I think it's clear that each time we, as society or as humans, experience a massive technological shift or advancement, it brings with it both opportunities and ways to make our life better or easier or move more smoothly and also challenges and dangers that are unknown to us in the development of that. And what I've heard from several of you today goes to the heart of this conversation, the need to balance the ethical, social, and legal implications with the technological advancement and the need to incorporate that from the beginning. So I want to address a couple of issues that Mx. Buolamwini—did I say that right?

Mx. BUOLAMWINI. Yes.

Ms. HORN. OK. And Ms. Whittaker especially have addressed in turn. The first is the incorporation of bias into AI systems that we are looking at more and more in our workplaces. This isn't just a fun technological exercise. So, Mx. Buolamwini, in your testimony you talked about inequity when it's put into the algorithms and also the need to incorporate social sciences.

So my question to you is how do we create a system that really addresses the groups that are most affected by this bias that could be built into the code and identifying it in the process? And then what would you suggest in terms of the ability to redress that, how to identify it and address it?

Mx. BUOLAMWINI. Absolutely. One thing I think we really need to focus on is how we define expertise, and who we consider the experts are generally not the people who are being impacted by these systems. So looking at ways we can actually work with marginalized communities during the design, development, deployment but also governance of these systems, so what—my community review panels that are part of the process, that are in the stakeholder meetings when you're doing things like algorithmic impact assessments and so forth, how do we actually bring people in.

This is also why I suggested the public interest technology clinics, right, because you're asking about how do we get to redress? Well, you don't necessarily know how to redress the issue you never saw, right? If you are denied the job, you don't know. And so there needs to be a way where we actually give people ways of reporting or connecting.

At the Algorithmic Justice League something we do is we have "bias in the wild" stories. This is how I began to learn about HireVue, which uses facial analysis and verbal and nonverbal cues to inform emotional engagement or problem-solving style. We got this notification from somebody who had interviewed at a large tech company and only after the fact found out that AI was used in the system in the first place. This is something I've also asked the FTC (Federal Trade Commission) about in terms of who do you go to when something like this happens?

Ms. HORN. Thank you very much. And, Ms. Whittaker, I want to turn to you. Several of the things that you have raised are concerning in a number of ways. And it strikes me that we're going to have to address this in a technological and social sciences setting but also as a legislative body and a Congress, setting some parameters around this that allow the development but also do our best to anticipate and guard for the problems, as you've mentioned.

So my question to you is, what would you suggest the role or some potential solutions that Congress could consider to take into account the challenges in workplace use of AI?

Ms. WHITTAKER. I want to emphasize my agreement with Mx. Buolamwini's answer. I will also point to the AI Now Institute's Algorithmic Impact Assessment Framework, which provides a multistep process for governance. The first step involves reviewing the components that go into creating a given AI system: examining what data informs the system, how the system designed, and what incentives are driving the creation and deployment of the system. The second involves examining the context where the system is slated to be deployed, for instance examining a workplace algorithm to understand whether it's being used to extract more profit, whether it's being designed in ways that protect labor rights, and asking how we measure and assess such things. And the third and critical step is engaging with the communities on the ground, who will bear the consequences of exploitative and biased systems. These are the people who will ultimately know how a given system is working in practice. Engineers in a Silicon Valley office aren't going to have this information. They don't build these systems to collect such data. So it's imperative that oversight involve both technical and policy expertise, and on-the-ground expertise. And recognize that the experience of those on the ground is often more important than the theories and assumptions of those who design and deploy these systems.

Ms. HORN. Thank you. My time is expired. I yield back.

Chairwoman JOHNSON. Thank you very much. Ms. Stevens.

Ms. STEVENS. Thank you, Madam Chair. Artificial intelligence, societal and ethical implications, likely the most important hearing taking place in this body today with profound implications on our future and obviously our present-day reality. Likely, the time we've allotted for this hearing is not enough. In fact, it might just be the beginning.

We've referenced it before, our proverb behind us, "Where there is no vision, the people will perish." And this is certainly an area where we need profound vision, a push toward the implications. And something that Mx. Buolamwini's statement in your testimony jumped out at me, which is that we have arrived overconfident and underprepared for artificial intelligence. And so I was wondering if each one of our panelists could talk about how we—not just as legislators are overconfident—in fact, I just think we're behind—but how we are underprepared. Thank you.

Ms. WHITTAKER. Well, I think one of the reasons we're overconfident is, as I said in my opening statement, that a lot of what we learn about AI is marketing from companies who want to sell it to us. This kind of marketing creates a lot of hype, which manifests in claims that AI can solve complex social problems, that its use

can produce almost magical efficiencies, that it can diagnose and even cure disease. And on and on.

But we're unprepared to examine and validate these systems against these claims. We have no established, public mechanism for ensuring that this tech actually does what the companies selling it say it does. For the past two decades the tech industry has been allowed to basically regulate itself. We've allowed those in the business of selling technology to own the future, assuming that what's good for the tech industry is good for the future. And it's clear that this needs to end.

In our 2018 annual report, AI Now recommended that truth in advertising laws be applied to AI technologies. All claims about AI's capabilities need to be validated and proven, and if you make a claim that can't be backed up, there will be penalties. The fact that such regulation would fundamentally change the way in which AI is designed and deployed should tell us something about how urgently it's needed.

Mr. CLARK. We're overconfident when it comes to believing these systems are repeatable and reliable. And as the testimonies have shown, that's repeatable for some, reliable for some. That's an area where people typically get stuff wrong.

As a society, we're underprepared because we're under-oriented. We don't know where this technology is going. We don't have granular data on how it's being developed. And the data that we do have is born out of industry, which has its own biases, so we need to build systems in government to let us measure, assess, and forecast for this technology.

Mx. BUOLAMWINI. First, I want to attribute Cathy O'Neil for we've arrived in the age of automation overconfident. I added underprepared because of all of the issues that I was seeing, and I do think part of the overconfidence is the assumption that good intentions will lead to a better outcome. And so oftentimes, I hear people saying, well, we want to use AI for good. And I ask do we even have good AI to begin with or are we sending parachutes with holes?

When it comes to being underprepared, so much reliance on data is part of why I use the term data is destiny, right? And if our data is reflecting current power shadows, current inequalities, we're destined to fail those who have already been marginalized.

Dr. TOURASSI. So what we covered today was very nicely the hope, the hype, and the hard truth of AI. We covered every aspect. And actually this is not new. The AI technologies that existed in the 1990s, they went through the same wave. What's different now is that we're moving a lot more—a lot faster because of access to data and access to computer resources. And there is no doubt that we will produce code much faster than we can produce regulations and policies. This is the reality.

Therefore, I believe that investments, strategic investments in R&D so that we can consistently and continuously benchmark datasets that are available for development of AI technology to capture biases to the extent that we can foresee these biases and continue to—continuously benchmark AI technology not only from the point of deployment but as a quality control throughout its lifetime, that needs to be part of our approach to the problem.

Ms. STEVENS. Well, thank you so much. And for the record, I just wanted to make note that earlier this year in this 116th Congress, I had the privilege of joining my colleague from Michigan, Congresswoman Brenda Lawrence, and our other colleague, Congressman Ro Khanna, to introduce H.R. 153, which supports the development of guidelines for the ethical development of artificial intelligence. So it's a resolution, but it's a step in that direction.

And certainly as this Committee continues to work with the National Institute of Standards and Technology and all of your fabulous expertise, we'll hopefully get to a good place. Thank you.

I yield back, Madam Chair.

Chairwoman JOHNSON. Thank you very much.

That concludes our questioning period. And I want to remind our witnesses that the record will remain open for 2 weeks for any additional statements from you or Members or any additional questions of the Committee.

The witnesses are now excused. I thank you profoundly for being here today. And the hearing is adjourned.

[Whereupon, at 12:03 p.m., the Committee was adjourned.]

Appendix I

Answers to Post-Hearing Questions

Answers to Post-Hearing Questions

Responses by Ms. Meredith Whittaker

Answers to questions from Congresswoman Haley Stevens:

1. There are a number of groups currently engaged in technical research into fair, transparent, and explainable AI systems. These include many of the people who publish at the ACM FAT* conference. Many of those engaged in this work are based in industry, and have made significant contributions. However, it is important to note that technical definitions of fairness *alone* will never be sufficient to address the issues of bias and harm that the use of AI systems in sensitive social contexts raise. In fact, a focus on narrow technical definitions without broader social awareness could serve to distract from more pervasive issues, which have less to do with the way a given AI system is calibrated in the lab, and more about the power asymmetries and historical injustices that exist within the domain where a given system is used. T

2. There are a number of competing definitions of "fairness," and as you note, the principles and ethics documents currently popular within the AI industry don't clearly define their use of this and similar terms. While a number of objectively measurable definitions do exist, these focus primarily on ensuring that an AI system produces parity, e.g. making sure that a given system treats members of protected classes the same as everyone else. These definitions are helpful in identifying instances where a system is definitively biased. However, they *are not sufficient* to guarantee that a system is not biased, or will not otherwise cause harm once it is deployed. In other words, technical standards and mathematical definitions are necessary, but entirely insufficient to ensure that these systems are used in ways that comport with justice and our democratic values.

   Any standard that is developed will need to be informed by perspectives from beyond the technical fields, and will need to focus on qualitative methods for monitoring, overseeing, and assessing the use of such systems within complex social contexts. Members of communities most at risk of harm will need to be at the table during such standardizing processes, as will social scientists, historians, and others with domain-specific expertise. A one-size-fits-all AI fairness standard is not possible, since AI systems interact with many separate domains, each of which engage different risks and opportunities. A sector-specific approach will be necessary, that focuses less on the system, and more on the histories and context of use.

Answer to the question from Congressman Roger Marshall:

1. I would need much more information to adequately answer this question. Including information on the populations from which the DOD data was collected, when these data were collected, the diagnostic methods by which pathology was/is identified in the lead-up to creating such data, in addition to information about how such a system would be tested and validated, whether (and by whom) it would ultimately be commercialized, and how it might be used (and on whom) in practice.

   I would note that preserving the privacy of the patients whose data is captured in the DOD's data may be possible. But creating a system to help diagnose future pathologies, based on this DOD data, also raises issues of privacy. Specifically the privacy of patients whose conditions may be detected by such a system in the future. This is particularly important in the context of the country's for-profit healthcare system, which ties access to medical care to employment, and which results in those with preexisting conditions often being unable to access care, making disclosure and diagnosis risky.

Answers to questions from Congressman Pete Olson:

1. This question presumes that we have a clear, uncontested way of measuring which country is at the "top" of AI research. We do not. Many of the metrics that are invoked when people discuss the "AI arms race" between the U.S., China, and Russia are extremely rough, and tell us little. These include: counting the number of AI researchers, or the number of startups, or the number of conference papers submitted by researchers in a given country, etc.. Such metrics do not account for urgent concerns, like whether the AI being produced actually works; whether it encodes bias or errors; whether it relies on precarious labor and other forms of exploitation in order to (appear to) function; and whether it comports with fundamental values, like democracy and liberty. Racing to "be the best at AI" without considering these is a race to the bottom.

2. A diverse Advisory Committee such as the one you and your colleagues outlined in the FUTURE of AI Act could certainly be helpful. Ensuring that a diverse array of experts are empowered to oversee the development, deployment, and assessment of AI, and to hold those who profit from biased or exploitative systems accountable, is long overdue. To be truly effective, such a Committee would need to include ample representation from communities who have endured historical discrimination, and the risks that AI presents for these communities must be at the center of the Committee's focus.

Responses by Mr. Jack Clark

HOUSE COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

Questions for the Record to:
Mr. Jack Clark
Policy Director
OpenAI
Submitted by Congresswoman Haley Stevens

1. There are calls for more technical research into fair, transparent, and explainable AI systems. Who might carry out this type of research? Is there a role for the private sector in contributing to this research or is it best left entirely to public funding? What level of investment is appropriate relative to our overall Federal investment in AI?

Answer: I think government has an important role to play in facilitating technical research into fair, transparent, and explainable AI systems. Specifically, government should facilitate greater conversation between industry, government, and academic stakeholders about technical research into these critical issues, potentially via hosting joint-workshops that draw on the federal government's research investments (for instance, the national lab system) as well as its ability to catalyze the development of standards via agencies such as NIST.

One affordable and effective way to catalyze greater industry investment into these types of research is for government to develop challenges that target these areas, modeled on the robotics and self-driving car challenges hosted by DARPA in recent years. For instance, in seeking to develop standards for ethical AI, NIST might conceive of technical challenges which could be run where, for instance, contestants need to develop an AI system that achieves a pre-agreed-upon explainability threshold for its decisions, or a method for analyzing datasets for traits that could lead to unfair outcomes.

The government should significantly increase its investments here. Given that the total non-DoD, non-DARPA AI R&D across the federal government was $973.5 million in 2020, according to the NITRD "Supplement to the President's FY2020 Budget" report, it would seem prudent to more than triple NIST's requested budget (so, an additional investment of greater than $40 million) to allow it to facilitate multiple studies, convenings, and targeted research projects here. We should also pursue a large range of targeted technical grants via NSF, which should be on the range of a net increase in grants of at least $100 million.

2. What do we mean by "fairness" when we talk about AI? Or trustworthiness, or explainability or the other characteristics called out in the dozens of AI principles documents being developed by industry and nations. Are these system characteristics that can be objectively measured and standardized, for example through the NIST standards development process?
    - If so, who needs to be at the table to develop those standards?
    - If not, are there other ways to address these factors early in the process, before it becomes a legal matter after a system has been deployed and demonstrated harm.

Answer: Terms like "fairness", "trustworthiness", "explainability", "safety", among others that show up in various principles documents, are in the process of being defined as a consequence of research within AI. Many of these terms are likely to eventually morph into standards. E.g, today we are beginning to develop some notions for how we can talk about bias with regard to AI systems and it's likely that increasingly precise definitions of bias will begin to be used in the scientific literature when assessing the "fairness" of a given AI system. Once there is a critical mass of research which has standardized around the same working definition of bias, then it would be a good time for the government to help facilitate the creation of a standard.

In terms of who should develop those standards, I think we should draw from agencies across government (coordinated via NIST), along with representatives from industry, academia, and - most importantly - communities potentially impacted by the standards being discussed. For instance, should the government seek to develop a standard for "unbiased facial recognition", it would be critical to consult with people in the communities where such facial recognition systems might eventually be deployed.

What I've described above should work as a general approach for the creation of standards in AI to help society capture the upside and minimize the downsides of AI progress: government, ideally via NIST, should carefully monitor the research literature relevant to a given AI term. Once it becomes clear researchers are starting to converge around a set of standard terms - potentially facilitated by additional written responses to various RFIs that NIST could send out - then NIST should facilitate a series of convenings between government, industry, academia, and potentially affected communities, to develop standards.

In addition to the above, government can invest in mechanisms to draw attention to these critical issues, thus giving academia and industry a greater incentive to work on them. Government could draw attention by, for instance, having organizations such as the Congressional Research Service or Government Accountability Office conduct studies relating to the above terms, producing objective information to catalyze public debates. Additionally, if government were to invest more in its own ability to study these issues - for instance, via a re-funded and re-established Office of Technology Assessment (OTA) - then it would have a greater ability to conduct research into the subject, increasing its ability to come up with efficient and effective legislative interventions.

HOUSE COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

Questions for the Record to:
Mr. Jack Clark
Policy Director
OpenAI
Submitted by Congressman Roger Marshall

1. Artificial intelligence holds a lot of promise for detecting and treating disease. For example, a team from Harvard Medical School has demonstrated that AI Deep Learning tools can help reduce the error rate associated with the diagnosis of breast cancer by 85 percent.

   AI systems need to be trained to do this kind of analysis. That means these systems need diverse training data. And it turns out the DoD joint pathology center holds the largest collection of pathology slides in the world - over 34 million unique pathology samples. Digitizing these slides and providing that gigantic dataset to the research community would be a boon for researchers that are developing AI for disease detection.

   Such an effort is even aligned with the President's Executive Order on AI, which charges agencies to "review their Federal data and models to identify opportunities to increase access and use by the greater non-Federal AI research community in a manner that benefits that community, while protecting safety, security, privacy, and confidentiality."

   Are you familiar with this issue, and do you think it would be a worthwhile investment for DoD to digitize these slides and provide them to the research community, doing it in a way that protects the privacy of the patients?

Answer: While OpenAI does not work directly in the field of medicine and AI, we are familiar with research in this domain - much of which relies on the availability of detailed medical data. Therefore, the digitization of such a large repository of pathology slides would no doubt prove to be useful to AI researchers, and digitizing this resource could unlock innovations from industry, academia, and government researchers who might study the data. Of course, efforts should be taken to protect the privacy of individuals' whose medical data is contained in the resulting dataset.

HOUSE COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

Questions for the Record to:
Mr. Jack Clark
Policy Director
OpenAI
Submitted by Congressman Pete Olson

1. Vladimir Putin has claimed "The nation that leads in AI will be the ruler of the world," and China has their "Made in China 2025" initiative emphasizing the need to a leader in the Technology. The good news is the United States leads the world in the number of top AI researchers. However, while the top research is still happing in the United States, we have seen a considerable uptick from China and others investment in the field. My question to the whole panel is how do we make sure the United States stays as the top researcher in AI?

Answer: There are two key things the United States needs to do to stay at the forefront of AI research: 1) Optimize for immigration of high-potential AI talent via ensuring we are the world's most attractive company for the world's smartest researchers. We should be making sure we can facilitate the migration of high potential researchers to America to work on technologies like AI and should avoid implementing strategies that would reduce immigration in this critical area. Additionally, we should significantly increase our funding for AI research, on the order of more than tripling existing NSF funding streams, so that we can maintain a flourishing US academic AI research sector.

2. During the 115th Congress, Rep. John Delaney and I along with Senators Cantwell and Young introduced legislation which would establish a diverse Federal Advisory Committee under the Department of Commerce to look at and review U.S. Competitiveness, Workforce Changes, Educational needs and Ethics as well as a list of many other issues within the AI fields. Do you believe a diverse Advisory Committee such as the one I've outlined would be helpful in working through many of these tough issues?

Answer: Yes, a diverse federal advisory committee would help us think about these issues.

3. Many companies have taken a proactive approach in making sure their algorithms are "Ethical". For example Axon recently put together the first AI Ethical Board to make their use of AI transparent and ethical. How can federally-funded basic research and standards development help the private sector develop ethical, safe and secure artificial intelligence and machine learning systems?
    a. How can we use public-private partnerships to advance AI for the public benefit of the U.S.?

Answer: I think federally-funded standards development organizations, like NIST, are one of the best levers we have for encouraging the development of ethical, safe and secure AI systems. Specifically, NIST should look to convene participants from academia, government, and the private sector around areas of AI development like 'fairness' and 'trustworthiness', then work with these participants to see if any of these areas can be developed further to aid the creation of standards that would facilitate further development and commercialization.

On the question of public-private partnerships, specifically, I think it would be very useful for public organizations to host competitions on areas central to their work. For instance, we could imagine the National Institutes of Health hosting a variety of competitions meant to catalyze development of AI research systems for frontier problems, and could encourage participation of the private sector in these competitions.

Responses by Mx. Joy Buolamwini

July 29th Responses Regarding:

United States House Committee on Science, Space and Technology

June 26, 2019

Hearing on

Artificial Intelligence: Societal and Ethical Implications

Witness

Joy Buolamwini
Founder, Algorithmic Justice League

Masters in Media Arts and Sciences, 2017, Massachusetts Institute of Technology
MSc Education (Learning & Technology), 2014, Distinction, University of Oxford
BS Computer Science, 2012, Highest Honors, Georgia Institute of Technology

PhD *Pending*, MIT Media Lab

For additional information, please contact Joy Buolamwini at joy@ajlunited.org

Dear Chairwoman Eddie Bernice Johnson,

Thank you again for the opportunity to testify on June 26th regarding the *Societal and Ethical Implications of Artificial Intelligence* and the invitation to address follow on questions from committee members as well as submit updates for the transcript and testimony.

Question Responses For:

- **Congresswoman Haley Stevens**
  - ***Who might carry out <u>technical</u> research on fair, transparent, and explainable AI?***
    - **We need to make sure that research on AI is interdisciplinary because a solely technical approach will not be adequate for addressing the sociotechnical implications of AI systems.** In addition to supporting researchers in tradition STEM domains, we need to make sure we are including researchers in the social sciences and humanities whose expertise on social, historical, and political factors that shape how AI impacts society is of critical importance. We need to also bring in experiential experts, people who are not necessarily based in traditional research venues but have frontline experience that can guide research on fair and equitable AI grounded in real-world experiences.
    - **Congress can promote deeper collaborations between AI researchers and organizations that work most closely with communities that are most harmed by algorithmic inequality**. Fund university/community partnerships both to study AI harms on marginalized groups, and also to do participatory design of AI that is rooted in the needs of marginalized communities. Such collaboration will give a much better contextual understanding of the impact of AI on society, and more importantly enable those who are most impacted by AI harms to be part of the process of counteracting these harms.

  - ***Is there a role for the public sector to play or is it best left to public funding?***
    - **Research that is focused on fairness, transparency, and accountability in AI systems can pose a conflict of interest when that research can compromise company profitability.** For example, Amazon has publicly attacked research showing gender and skin-type bias in its controversial Amazon Rekognition AI product. Still, federally funded studies from the National Institute of Standards and Technology and the Maryland Biometrics Testing Facility show that for leading commercial facial analysis systems gender and skin-type do impact

system accuracy and performance.[1] This case shows the importance of third-party evaluation.

- **One way the public sector can support research in this domain is by making their AI systems available for testing by verified organizations and academic institutions.** Further, congress can make a provision that for any AI vendor seeking government contracts in a sensitive domain like law enforcement, the vendors proposed systems must be submitted to relevant national benchmarks. In the domain of facial analysis, companies like Microsoft and NEC have voluntarily submitted their models, but other vendors who are selling in sensitive domains have not. By requiring those who hope to secure government contracts submit models for third-party evaluation to inform research we can increase transparency. Government agencies and officials will also have the needed information for decision-making around the technical suitability of AI models.

- ***What level of investment is appropriate relative to Federal investment in AI?*** If spending is a reflection of priorities, ensuring the AI systems we build do not amplify racism, sexism, ableism, and other harmful forms of intersecting discrimination should be at the top of the list. We should devote a similar amount of resources to anticipating and mitigating the harms of AI as we do to advancing technical capacity.

- ***What do we mean by fairness in AI?***
  - **There is no singular definition of fairness in AI and often the definitions chosen are a reflection of the priorities and methodologies available to researchers asking the questions.** A computer scientist may attempt to reduce fairness down to optimization problems and formal equations. Yet a system deemed fair because based on a selected metric, it performs equally well across all groups of interest can still be abused. For example, facial recognition technology can be improved, but if is deployed to enable mass surveillance or integrated into lethal autonomous weapons, technical definitions of fairness quickly give way to ethical consideration concerning what technological deployments are appropriate for a fair, just, and free society. As noted in answering related questions, an interdisciplinary approach to studying fairness is needed and meaningful transparency facilitate debates on fairness.

[1] https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/

- ***Can characteristics like fairness be objectively measured and who might carry out evaluations of these kinds of characteristics in AI systems?***
  - **Organizations like the National Institute of Standards and Technology(NIST) are currently positioned to help provide transparency into the technical performance of AI systems which does not necessarily measure fairness as notions of fairness do not exist in a vacuum.** When characteristics like fairness or trustworthiness are framed as relational instead of stand alone attributes, it becomes clear that there will be no singular objective measurement. Context will always matter. Values are key, yet highly contested and evolving in ways that reflect the priorities of those who hold power in society. What is a fair way to distribute resources given existing factors like past inequalities and future aspirations? Standards bodies like NIST can provide important mechanisms for transparency. For example, with facial recognition technology by requiring that NIST conduct performance tests on demographic effects of factors like race and gender and phenotypic factors like skin types, decision makers can have the requisite technical information to inform decisions about the use of the technology that must also take into account social considerations. Even if a facial recognition system is deemed to be technically accurate, public scrutiny and deliberation is still needed before the technology is deployed. Even if the technology is deemed suitable for deployment in some cases, mechanisms for oversight, transparency, and meaningful redress for harms are necessary. **Addressing issues of fairness in AI is like hygiene. To be effective, it must be continuous and adaptive.**

- **Congressman Roger Marshall**
  - ***The President's executive order encourages a review of federal data and modules to find opportunities for access to the greater non-Federal research community. The DOD currently has over 34 million unique pathology samples taken from around the world. Would it be a worthwhile investment to digitize these samples?***
    - **The history of medical research has been plagued with issues of bias and major ethical oversights, so ensuring that we learn from the mistakes of the past requires addressing key questions about the data provenance of the pathology samples and potential distribution.**
      - How was the current data collected? Was consent given for a specific use case or follow on uses? In cases where consent was given for a specific use case, are there existing mechanisms to enable wider uses?

- Were subjects compensated for data either monetarily or in some other way? If not are there any feasible mechanisms for compensation? Short of or in addition to compensation, will an acknowledgement of how the data was collected and from which regions of the world be made known? How can we avoid exploitation? For example, HeLa cells which were extracted from Henrietta Lacks have enabled significant progress in scientific research, yet as chronicled in the *The Immortal Life of Henrietta Lacks* , she nor her family received substantive remuneration. We can do better in the age of AI.
- If data is made freely available will there be provisions that the research and findings from data also be made freely available to the general public. Currently, even though federal funds support medical research, practices of the academic publishing industry limit access to knowledge made possible through tax dollars. Without access to a university with costly subscriptions to specific research journals, knowledge is sequestered and limited to the well-resourced. Will researchers be incentivized to use open access platforms like PloS so more people can learn from knowledge enabled by the dataset?
- What mechanism will ensure the availability of the data? During the 2019 government shut down, I was dismayed and surprised to find I could not access certain government datasets (IJB-C) or websites due to the shut down. Some research projects that relied on access to government data were compromised, delayed, or cancelled. What redundancy measures will be put into place so research is uninterrupted even if the government is at a standstill?

- **Congressman Pete Olsen**
  - ***How do we make sure that the US stays on top in AI research?***
    - **In the June 2018 Congressional Hearing "AI: With Great Power Comes Great Responsibility" written testimony,[2] Professor Jaime Carbonell pointed to the investment in AI being made in China.**
      - *Excerpt-* US dominance in AI is being challenged like never before. Many countries are striving hard to improve their AI know-how, work-force, and industry, including China, Russia, Korea, Japan, Germany, the UK and India. Consider China, for instance, which has made AI a national priority. On May 4, 2018 CNBC reported "China is determined to steal the AI crown from the US and nothing, not even a trade war will stop it. China's 2030 plan envisions building a $1 trillion AI industry." Wired Magazine

[2] https://science.house.gov/imo/media/doc/Carbonell%20Testimony.PDF

reports: "China will be the world's dominant player in artificial intelligence by 2030. This isn't a prediction by a researcher or academic, it's government policy from Beijing." Whereas these statements may be on more alarmist than reliable predictions, they clearly indicate Chinese intent. China's national priority is AI preeminence. Even General Secretary Xi Jinping is reported to have AI books on his shelf. It is difficult to estimate the very substantial level of AI funding in China, but there some components include: 1) The city of Tianjin is committing $5 billion to support the new AI industrial park. 2) The Feb 20, 2018 statement in the Financial Times saying "Last year almost half the global investment into AI startups went to China, up from a mere 11.3 per cent slice in 9 2016", 3) On June 22, 2018 the South China Morning Post reported: "China's Ministry of Science and Technology has funded at least eight AI-related research projects over the past six months to the tune of 2.73 billion yuan (US$430 million) from the central government budget" and "The China Academy of Sciences (CAS) which has over 300 labs and four national research centres, received over 2.7 billion yuan for its 11 fundamental science projects last year, although it's unclear how many of these are directly-AI related." China has already far surpassed the US in terms of patents granted for AI technologies, according to Quartz, May 2, 2018.

To be an AI research leader, the US has to devote substantial resources to R&D.

- **The US also has the opportunity to be a leader no just in the technical development of AI but the ethical development of AI.** Instead of asking how does the US win the contrived AI race as is done in some headlines, let's ask "How can all of humanity thrive in the age of automation?" If "winning" in AI means compromising civil liberties and rights, in the end such an accomplishment will only be a pyrrhic victory.

- ***Do you support the establishment of a diverse Federal Advisory Committee under the Department of Commerce to review US Competitiveness, Workforce Changes, Educational Needs and Challenges as it relates to AI?***
  - **Ensuring Congress has access to requisite expertise is an essential component to equip lawmakers with the insights needed to inform decision making on one of the most transformative technological developments of humanity.** Making sure diverse expertise includes experiential expertise will be crucial.

Written Testimony Update
Misstatement: "Malissa Nielsen, a 24-year-old babysitter who stated she had nothing to hide, submitted her social media information and was surprised to find she was flagged, losing her job in the process." -page 11

Correction: "Malissa Nielsen, a 24-year-old babysitter who stated she had nothing to hide, submitted her social media information and was surprised to find she was given low ratings and was unaware these rating could impact her only source of income.[3]

[3] Drew Harrell, "Wanted: The 'perfect babysitter.' Must pass AI scan for respect and attitude.", (November 2018) https://www.washingtonpost.com/technology/2018/11/16/wanted-perfect-babysitter-must-pass-ai-scan-respect-attitude

Responses by Dr. Georgia Tourassi

HSST Hearing: Artificial Intelligence: Societal and Ethical Implications

Responses to Additional Questions

Georgia Tourassi

Director, Health Data Sciences Institute

Distinguished Scientist

Oak Ridge National Laboratory

July 26, 2019

Congresswoman Haley Stevens:

1. Congresswoman Stevens – I believe that this type of research is of paramount importance and must be inclusive of both the public and private sectors. We are all impacted by the advancements in AI and for the fair, transparent, explainable, and ethical considerations to be uniformly accepted we have to have buy in from both sides. The ethical and social implications of AI differ across applications domains and those domains must have a voice when it comes to the metrics by which we, as a society, grade this endeavor. Since the private sector is primarily driven by applications with financial incentives for their stock holders, public funding and federal investments at national labs and academia are necessary to balance the imperative and avoid application gaps which can avertedly lead to another form of bias or discrimination. There needs to be a significant level of federal involvement in developing a strategy and R&D investments in the technical aspects of fair, transparent, and explainable AI systems.

   I cannot speak directly what will be an appropriate level of overall Federal investment. I believe this to be a new frontier and will require a renewed look at the federal landscape of research investments. History has shown us that this will be a significant undertaking that must be sustained for years, if not decades. The Human Genome Project could serve as in informative example. The project, which was initiated by DOE, was carried out for 13 years and a total funding of roughly $3.6B federal investment through its completion in 2003. From its earliest stages, it included a program to address the ethical, legal, and social implications (ELSI) of genomic research. The ELSI program was funded at 3 to 5% of the total project budget and the National Human Genome Research Institute continues to fund ELSI research. Any initiative in AI will need this component, but holistically it needs to be prepared to endure a similar timeframe if not

a never-ending commitment as the field of computational science and artificial intelligence will continue to expand and impact our nation.

2. NIST is ideally suited to lead the effort of standards development for assessing fairness, explainability, and transparency of AI technology. The standards and related metrics will differ across application domains and will need to be consistently revised and updated as AI technology evolves. NIST will need a strong partner to ensure proper and honest implementation of the standards as well as pre-deployment quality assessment and continuous post-deployment quality monitoring of AI technologies.

Congressman Roger Marshall:

1. Congressman Marshall - I agree. It is worthwhile for DoD to invest in digitizing the vast volume of pathology slides it has available to enable research and development of large-scale AI solutions in this space. In medical imaging, automated analysis of histopathological images is a well-known grand challenge due to the sheer size and complexity of these images. AI has a lot to offer in this space, as documented by small scale single-institution studies. In addition, there needs to be additional investments in establishing gold standard for each pathology slide so that developed AI solutions can be properly evaluated and benchmarked.

   Beyond the DoD pathology sides, I believe such an undertaking to be a worthwhile investment across all government owned and maintained datasets. A concerted effort to curate and make available government data for research while protecting patient privacy would have profound positive implications for the scientific community and national competitiveness. However, we must also remain cognizant that there needs to be an equivalent consideration for infrastructure to handle and utilize the vast data assets. Access to immense computing power to develop such solutions will also be critical. Co-locating these federal data assets in a secure environment with the appropriate computing resources will help accelerate advances.

Congressman Pete Olsen:

1. Congressman Olson, I also believe that there are profound benefits to being the global leader in artificial intelligence. To maintain our leadership in the field of AI, I would recommend that the United States consider a multi-faceted approach that leverages not only our research infrastructure, but also educational infrastructure to ready a new generation of interdisciplinary scientists. As a nation we must maintain and increase the level of R&D investments in AI across federal agencies and develop a roadmap for better federal coordination of resources and interagency strategic partnerships, particularly with respect to federal data assets. The data will be the key enabler of this new paradigm and considerable resources should be considered to enhance the readiness and quality of the data (cleaning, curation, etc.).

   Additionally, in my opinion there exists an opportunity to influence STEM education and support workforce development via a life-long learning curriculum with re-skilling opportunities to maintain agility in the ever-shifting landscape of computational science

and artificial intelligence. Additionally, concentrated STEM development across underrepresented communities such as Historically Black Colleges and Universities is essential to ensure a diverse and technically thriving US workforce and development and deployment of socially responsible AI.

2. I believe this will be a critical step to maintain the nation's competitive advantage and leadership in the field of AI. Additionally, I would recommend that this be a multi-agency committee that can bring to the table various facets of our nation's government. This is an all-of-government challenge with different nuances represented by the missions of multiple agencies.

Appendix II

Additional Material for the Record

116TH CONGRESS
1ST SESSION

H. RES. 153

Supporting the development of guidelines for ethical development of artificial intelligence.

IN THE HOUSE OF REPRESENTATIVES

FEBRUARY 27, 2019

Mrs. LAWRENCE (for herself, Mr. KHANNA, Mr. SOTO, Mr. LIPINSKI, Mr. CRIST, Ms. STEVENS, Ms. KELLY of Illinois, Ms. DELBENE, and Ms. MENG) submitted the following resolution; which was referred to the Committee on Science, Space, and Technology

---

RESOLUTION

Supporting the development of guidelines for ethical development of artificial intelligence.

Whereas the field of artificial intelligence (AI) was initiated by a single question, "Can machines think?", has made significant advancement since the 1950s, and today touches every aspect of American society;

Whereas AI has demonstrated increasing competency in areas as diverse as image and speech recognition, autonomous driving, and the mastery of complex games;

Whereas AI has the potential to transform the economy and dramatically alter industries including health care, retail, finance, energy, transportation, law, education, and manufacturing over the coming years;

Whereas the development and use of AI has the potential to enhance wellbeing, foster economic growth, and improve care and services for many people;

Whereas the far-reaching societal impacts of AI necessitates its safe, responsible, and democratic development; and

Whereas the leaders of the G7 have committed to the Charlevoix Common Vision for the Future of Artificial Intelligence: Now, therefore, be it

1 *Resolved,* That the House of Representatives supports
2 the development of guidelines for the ethical development
3 of artificial intelligence (AI), in consultation with diverse
4 stakeholders, and consonant with the following aims of:
5 (1) Engagement among industry, government,
6 academia, and civil society.
7 (2) Transparency and explainability of AI sys-
8 tems, processes, and implications.
9 (3) Helping to empower women and underrep-
10 resented or marginalized populations.
11 (4) Information privacy and the protection of
12 one's personal data.
13 (5) Career opportunity to find meaningful work
14 and maintain a livelihood.
15 (6) Accountability and oversight for all auto-
16 mated decisionmaking.
17 (7) Lifelong learning in STEM, social sciences,
18 and humanities.

1 (8) Access and fairness regarding technological
2 services and benefits.
3 (9) Interdisciplinary research about AI that is
4 safe and beneficial.
5 (10) Safety, security, and control of AI systems
6 now and in the future.

○

○