U.S. PRIVACY AND CIVIL LIBERTIES OVERSIGHT BOARD

(PCLOB)

PUBLIC FORUM ON ARTIFICIAL INTELLIGENCE

Thursday, July 11, 2024

U.S. PRIVACY AND CIVIL LIBERTIES OVERSIGHT BOARD

A P P E A R A N C E S

SHARON BRADFORD FRANKLIN
Privacy and Civil Liberties Oversight Board Chair

BETH A. WILLIAMS
Board Member

TRAVIS LEBLANC
Board Member

EDWARD W. FELTEN
Board Member

MIKE ROUNDS
United States Senator, R-South Dakota

ALONDRA NELSON
Former White House Office of Science and Technology Policy Acting Director

DEAN SOULELES

Former Office of the Director of National Intelligence Chief Technology Advisor

ELHAM TABASSI
National Institute of Standards and Technology Senior Scientist

WILLIAM USHER
Special Competitive Studies Project Senior Intelligence Director

MIRANDA BOGEN
Center for Democracy and Technology AI Governance Lab Director

CLARE GARVIE
National Association of Criminal Defense Lawyers Counsel

JAMIL JAFFER
George Mason Law School National Security Institute Director

PETER WINN

Justice Department Acting Chief Privacy and Civil Liberties Officer

P R O C E E D I N G S

MS. FRANKLIN: Hello. I'm Sharon Bradford Franklin, chair of the Privacy and Civil Liberties Oversight Board. Together with my fellow board members, Ed Felten, Travis LeBlanc, and Beth Williams, I'd like to welcome you to today's public forum on the role of artificial intelligence and counterterrorism and related national security programs and the privacy and civil liberties issues associated with these uses of AI.

The uses of AI in all facets of our lives are rapidly and continually growing, as is the sophistication of these tools. These trends are raising a variety of questions for policymakers, ranging from overarching concerns like how to encourage American competitiveness in AI and what limit should be put on AI to avoid harmful outcomes. To more day-to-day and specific questions, such as how to prevent cheating by students who want to use ChatGPT to write their research papers.

But the Board's focus on AI is not motivated

simply by a desire to get in on the latest big tech graves. Rather, as our government incorporates AI tools into its efforts to protect the nation from terrorism, it is our role to ensure that those government strategies also protect individual rights and liberties. Yet the potential uses of AI are extensive, and we must be strategic in our oversight of the government's use of AI for counterterrorism purposes.

Even before most people became aware of generative AI and tools like ChatGPT, there's been a lot of research on the privacy and civil liberties risks posed by AI tools. As I expect, we will discuss further today, these range from reliance on training data that reflects and perpetuates patterns of historical discrimination, to bias and facial recognition systems that don't work as well for particular racial or other demographic groups, to AI-based decision making that lacks explainability or other due process safeguards.

On the other hand, AI offers enormous benefits. And I'm encouraged by the various efforts

to develop frameworks to address the risks posed by the uses of AI. In recent years, these have included the artificial intelligence ethics framework for the Intelligence Community, and the White House blueprint for an AI Bill of Rights.

Just last fall, the President issued a new executive order on the safe, secure and trustworthy development and use of artificial intelligence. And we expect the issuance of a National Security Memorandum, or NSM, on AI before the end of this month.

But how well is the government doing it translating these principles into action? What gaps remain in these frameworks? And how can the Board best focus our resources to conduct our oversight on the use of AI in counterterrorism?

So, today's public forum is designed to inform both our Board and the public as we work to scope and define our oversight of the government's use of AI in counterterrorism and related national security programs.

Before we turn to that conversation, I have

just a couple of notes on today's event. First, I want to thank our staff for all their tremendous work planning today's forum and making it possible for us to come together online. And then in terms of logistics. Today's forum of course is limited to matters that can be discussed in this unclassified public setting. Our format will include remarks from Senator Rounds, followed by two panels.

Today's event is being recorded, and the recording will be posted on our website. For each panel we will first hear brief opening statements from each panelist, then my fellow Board members and I will take turns asking questions of the panelists. And we will cycle through our order again as time permits.

So, we will begin our event today with some pre-recorded remarks from Senator Mike Rounds. He is co-chair of the Senate AI Caucus and a member of the bipartisan Senate AI Working Group. And he also sits on the Senate Select Committee on Intelligence.

So, over now to those remarks.

UNIDENTIFIED SPEAKER: My apologies. I will start video with audio.

MR. ROUNDS: Hi, I'm Senator Mike Rounds. Thank you for the opportunity to talk to you today. I know that the use of AI by the Intelligence Community for counterterrorism and other IC missions is a topic of particular interest to you.

Our Intelligence Community collects enormous amounts of multi-source data each day that the United States uses to support national security priorities and objectives to include counterterrorism. AI is capable of processing huge amounts of data, which in turn is being utilized to identify patterns of life and to analyze significant amounts of data in a very short period of time. Such identification and analysis can continue to enable the IC to more effectively and efficiently track suspected or known terrorists as well as terrorist financing activities.

For example, the IC's Project Maven uses an AI tool designed to process imagery and full motion video from unmanned systems and can systematically detect potential targets for collection. This will allow us to more efficiently identify and neutralize terrorists.

As a member of the Senate Select Committee on Intelligence, I look forward to further fostering these efforts. As our government matures these capabilities, it will be important to establish metrics to measure the performance and efficacy of these AI supported capabilities. Such performance metrics for the IC's use of AI could include measuring the speed of analyzing intelligence data sets across collection platforms and the breadth of resource utilization, as well as the depth of global collection, all balanced against protecting liberties and Fourth Amendment privacy protections.

I believe that measuring these kinds of performance metrics must be a part of the intelligence committee's oversight of the IC. As we foster new IC AI capabilities and performance metrics, we should also make certain that this new capability adheres to constitutional standards and privacy rights. I believe maintaining those standards will continue to be a key element of the Intelligence Committee's oversight role.

The Director of National Intelligence with

input from relevant departments and agencies, bears the responsibility for making sure that the IC tests and safeguards AI systems before deploying them. I should also point out that our nation will face AI-generated threats that not only include direct military threats in all five war fighting domains of air, land, sea, space, and cyberspace, but also include threats to our larger society. Perhaps most importantly, that includes threats to our critical infrastructure.

These threats come from nation states, terrorists, and criminal organizations. AI will be used to generate photo, audio, video, and other forgeries of elected officials or other public figures making incendiary comments or behaving inappropriately, the so-called deepfakes. Doing so could potentially erode public trust, negatively affect public discourse, and even potentially sway an election. Such AI-generated products could also be used to embarrass or blackmail elected officials or individuals with access to classified information.

A major concern I share with some of my

colleagues, particularly senators Young, Schumer, Heinrich, and they're all a group with whom I've worked extensively over the last year on AI policy, is the use of AI to generate novel biological threats.

Finally, offensive military use of AI is accelerating the pace of combat. This reduces decision-making timelines for the defender and reduces the opportunity to deter, or if deterrence fails, to defeat an attack. As one of the few members of the Senate who sits on both the Intelligence and Armed Services committees, addressing these AI-generated threats will continue to be a priority. Thank you again for the opportunity to talk to you today. I look forward to a continued dialogue with you in the challenging days ahead for our national security in the age of AI. Thank you.

MS. FRANKLIN: Okay. Apologies to our audience for the technical difficulties, but thank you to our IT team for making that work.

So, now, hopefully all of our panelists for the first panel will join us, turn their cameras on and welcome to you. Thank you for joining us. So,

for our first panel, we will hear from in alphabetical order, I believe, first Alondra Nelson, who is a former acting director of the White House Office of Science and Technology Policy or OSTP. Then Dean -- sorry, I'm going to mess up your name, Souleles, yes. former chief technology advisor for the Office of the Director of National Intelligence, then Elham Tabassi, senior scientist at the National Institute of Standards and Technology, or NIST, and then William Usher, senior director for intelligence at the Special Competitive Studies Project.

And so for each panelist will make, in order that I just went through, opening remarks up to 5 minutes and then we will turn to questioning from the board members. So, Alondra Nelson first. Thank you.

MS. NELSON: Good morning. Thank you, Chair Franklin, and members of the Privacy and Civil Liberties Oversight Board. Thank you for convening this critical public discussion on issues associated with the use of AI in the national security context. And I'm honored to be with this distinguished panel.

So, I'm a social science scholar and

researcher and policy adviser who spent 26 months serving in the leadership of the White House Office of Science and Technology Policy in the Biden-Harris administration. During my OSTP tenure, we stood up the National AI Initiative Office to coordinate AI policy across the whole of government. The National Science and Technology Council that OSTP administers on behalf of the President issued an updated list of critical and emerging technologies, the subset of advanced technologies that are potentially significant to U.S. national security. This list included not only many forms of artificial intelligence, but a number of other technologies that we often consider advanced in part because of their use of systems of data collection, analysis and dissemination that include forms of automation in whole -- in part or whole.

At OSTP and my time there, we also launched the National AI Research Resource Task Force, the recommendations of which led to a pilot program to democratize access to the data and compute required for responsible AI development. And we developed, as

Chair Franklin mentioned, the blueprint for an AI Bill of Rights, a cornerstone of Biden-Harris AI policy that distills best principles and practices for guiding the safe and responsible design, development and deployment of AI technologies.

In my past and current research, I also studied the social implications of science and technology -- of science and technology and related policy and research analysis issues. Across this work, I've come to appreciate that particular challenges that advanced AI presents to both national security, including counterterrorism especially to the -- especially acute regarding the preservation of our principles, norms, and practices we need to protect rights and liberties.

AI technologies, both so called Predictive AI and more recent generative AI, have expansive potential use in the national security context and do a lot of work to keep us safe, including intelligence data processing and research, strategic decision making with humans on the loop or in the loop as the case may be, transportation logistics, cybersecurity,

there's a growing use of drones, which we should probably discuss, targeting and simulation.

One of the examples of use for national defense or planetary defense, moreover, that I often like to talk about is in the space of outer space and international and space policy. You might be familiar with the double asteroid redirection test or the DART mission, which is part of U.S. national and planetary defense. It was designed and carried out to protect Earth from collision with an asteroid or another entity by moving an object out of its orbit and out of therefore a dangerous trajectory. NASA succeeded in this mission for the first time in late 2022. And this was made possible by years if AI-enabled calculation and autonomous simulation, more particularly the Small-body Maneuvering Autonomous Real Time Navigation algorithms or SMART Nav algorithms that allow scientists to predict the path of an asteroid, and then to plan the navigation of a spacecraft to collide with it, and place it on a non-harmful path and also not cause harm to the spacecraft.

Crucially important for national and planetary defense, therefore, are -- is something like the DART mission and also is critically important science for the volume of orbital debris, the satellite launches that grow every day, and the kind of geopolitics of space that's happening that poses new national security risks.

But I think our discussion today is no doubt about the implications of AI in the national security context prompted by the developments in advanced AI since November of 2022 when ChatGPT was released to the world and the emergence of these kinds of foundation models and what they mean for, as Senator Round suggested, the generation of text, of sound, and image that have been described as general purpose.

General purpose, that phrase lies -- herein lies the challenge that AI poses, both the opportunity and the challenge that AI poses for national security. For this new suite of technologies threaten to thicken the so called fog of war, that disorientation and uncertainty of situational awareness in the military theater, they threaten to thicken the fog of war to

brattle social effects across both civilian and military domains.

So, we might call this potential, the fog of advanced AI, right, and it has a few important facets for our discussion. One, that we are increasingly with advanced AI using inscrutable commercial AI software that can be transformed into many forms that are not fully known. Some of them are quite banal, and some of them might be dangerous, but we don't know.

Second and related. The black box that is often necessary for military and IC secrecy with these new inscrutable technologies is compounded and further obscured by an accuracy by biases in the technology and the training data, and by the fundamental weakness of inscrutable technology like generative AI that for many use cases works pretty well a lot of the time, but doesn't work entirely well all of the time.

The implications for one and two for the commercial software that can be used for both dangerous and banal uses, that compounds the black box of sometimes necessary military secrecy, means that

layered on to defense secrecy is this layer of black box technology that holds significant implications for national security effectiveness and also for public accountability.

The traditional notions of dual use technology are technologies that are intended for one purpose and that can have been discovered often to have an application for another use, one purpose being civilian, the other military.

A classic case emerging from chemical and biological research has been the development of, you know, bio weapons beginning in the early 20th century. And more recently, we had the development of massive explosive capabilities from the use of ammonium nitrate fertilizer and other chemicals combined that were widely available to carry out the Oklahoma City bombing.

This act of domestic terrorism is a perfect analogy for advanced AI and that many civilian and military applications can be made inherently out of the work -- out of generative AI. These can be both intended and unintended use cases.

For example, we might take the case of facial recognition technology. We know, for example, from reporting, as Chair Franklin mentioned, this is all widely known information that Clearview AI's facial recognition technology is being used in the Russia-Ukraine war, being used by Ukraine to identify deceased Russian soldiers. Clearview's AI systems are known to be built from scraping websites of civilian data, creating potential rights violations in a civilian context importing these into the theater of war.

Without public accountability, and there's -- these technologies are often -- also used for public security. So, this is not just one technology intended to use in one domain and used in another, what we face today is the circulation of these technologies back and forth across civilian and military domains simultaneously in ways that create new challenges for oversight boards like this one for policymakers who work both on the civilian and military sides and that raise tensions for democratic societies.

Facial recognition technology used domestically by police, including DataWorks Plus in Detroit has yielded numerous cases of misidentification that I bet have had high costs for people's lives, including for Robert Williams, an African American man arrested in front of his family for burglary he wasn't involved with.

MS. FRANKLIN: Thank you, if you could just wrap up your opening so we can move on to the other panelists and hopefully have more time for questions. Thank you.

MS. NELSON: Sure. Yeah, yeah, okay. So, to date the government has -- what is clear is that the US will need to develop new standards of practice and engagement that do not adhere to the technology not to AI but to the mission and values of the U.S. And this is because these technology, commercial technologies will have to be -- decisions about them will have to be shared not only across the IC, but across the Department of Commerce, FTC and other executive agencies. Public accountability has always been hard to accomplish regarding military uses of technology.

But this becomes more urgent in the context of general purpose dual use technologies.

With the introduction of advanced AI, we can no longer effectively or neatly separate civilian laws and regulations from military ones. War is often the best -- worst way to preserve a way of life and to use AI in a way that diminishes our basic values is not mission-aligned. Allied countries can work together to minimize abuse by reducing the circulation and dissemination of commercial AI technologies with export controls and sanctions.

But fundamentally an unregulated U.S. commercial AI technology industry with dual use general purpose technology increases national security risks. Fundamental regulation is needed. I know this is not the mandate or domain of authority for the board. However, the board can use its sphere of influence to see where the various responsible use of AI exist.

MS. FRANKLIN: Thank you. Thank you. I'm so sorry to interrupt you. But I do want to make sure we have time for everybody's opening, and then for the

questioning with the board. Thank you so much.

Okay, so we will next hear opening remarks from Dean Souleles.

MR. SOULELES: Perfect. Thank you so much. Thanks for convening this session. It's a very important session. My role in government and as a career technologist was often at the intersection of technology and management or mission, and to translate for the technologists what the actual mission is, and to translate for the mission what the technology is and what it is and what its limitations are. So, I kind of sat at that intersection in my time at the Office of Director of National Intelligence. And I want to talk a little bit about that, from that perspective.

AI is clearly very important to the counterterrorism mission but, as always, I'd like to start with defining our terms. So, when I speak of the counterterrorism mission, I'm speaking very simply of our mission to collect, analyze, and share actionable intelligence related to terrorism and to detect and disrupt those threats. So, within that

context, we need to look at what AI is and what it isn't.

And, in addition, at the Office of Director of National Intelligence, the National Counterterrorism Center, is responsible for maintaining the authoritative database of known and suspected terrorists. So, we got a big database of people. That's a identification issue. So, that's the ICCT mission. But what do we mean by AI? And if you have a conversation about AI and civil liberties, you better know what you're talking about. And that's not so easy to answer.

In the current environment, you could be excused for thinking that AI is synonymous with large language models and chatbots. If you haven't been deeply involved in technology, it appears that this technology came out of nowhere 2 years ago. Well, it didn't.

And it's now seemingly everywhere, it's pervasive. But this is the latest in a long, long line of machine intelligence tools that have become increasingly and more useful over the last decade.

And by the way, the DOD and the U.S. Intelligence Community have been using these tools for years, many, many years. This is just the latest in a set of technologies.

In 2012, Yann LeCun and Geoffrey Hinton demonstrated neural network based supervised machine learning was better than or equivalent to human in many cases. And that triggered this wave of technology that turned into the kinds of technologies that we are seeing today. I broadly classify AI tools into a bunch of buckets, there's a bunch of different taxonomies. But a useful one is to think about supervised machine learning. This is where we take large amounts of ground truth data, which is called training data, usually provided and curated by human experts into a machine classification system. That's how image recognition and facial recognition works.

And as we've heard, if you put the wrong date in, you're going to get the wrong conclusions out, it's going to have bias. Then there's unsupervised learning, which takes massive amounts of data and seeks to find patterns or connections in the data in a

way to make it more useful.

Again, it's only going to produce relevant insights based on the data that has been fed. And we make a bias decision every time we choose what to include, or what not what not to include in those systems. And what questions to ask of those systems.

Another kind of AI is reinforcement learning. And this is a set of AI technologies where the system learns how to behave in ways that increase reward, they call it -- mathematicians call it a reward function, by interacting with the environment. In other words, the AI gets it right, you increase the reward, which is a numerical number, if it gets it wrong, you decrease it. And you run these things many tens of thousands or millions of times and that's how you get a computer that can beat the best players at Chess and Go with this idea of reinforcement learning. That's not as intended, that's not dependent on data, it's really dependent on a set of rules. But you can bias that system however you like, by choosing the rules in which you wish to train it.

Deep learning underlies all that. And it's a

set of technologies that work across all the areas I just talked about. It uses large quantities of data to figure out how to do complex things, searching through combinations of ways that best describe the data. So, that's kind of the context of this. And all of those things, all the things that we think of as AI are one of those sorts of things. And they're basically computer decision making, computer search, advanced decision making advanced analytics, there's all kinds of ways you can describe it. But at the end of the day, they are mathematical models that help us make conclusions about data.

We may have done ourselves a disservice by personifying things like ChatGPT and having it speak in human terms. These are not humans, they are not brains, they do not think, they do not have hopes and dreams, you turn them off, and they don't -- they go to sleep and they don't come back. We need to be aware of what it is we are talking about. So, when we set up the AI strategy for the IC, I felt it was very important that we address things like what are the risks, and we know that AI can learn the wrong thing,

if it's given the wrong data, we know it can do the wrong thing. And worse, it can do it with confidence. It will always give you an answer.

And you need to be aware as a human analyst, that the fact that it gives you an answer isn't the fact that it is correct. And it can even reveal the wrong thing. So, in the context of national security, if our models, our classified black box models leak out into the world, we know that as analysts we can analyze those models and learn what training data that we were training them on. So, these are risks. Now there are a huge number of other things that we can talk about, and Alondra mentioned many of those. I won't repeat them. But I'm happy to happy to take any of your questions in the question round.

MS. FRANKLIN: Thank you. Okay, so we will next hear from Elham Tabassi.

MS. TABASSI: Good morning, everyone. Grateful for the invitation, chair Franklin, member of the Boards. I'm delighted to be here among distinguished speakers. Always difficult to follow Dr. Nelson and also Dean, but I try to do my best.

Again, thanks for the opportunity to come and talk about some of the things that we have done in the space of AI and AI risk management.

For some of the audience that may not know NIST, National Institute of Standards and Technology, we are a measurement science agency. NIST is a nonregulatory agency under Department of Commerce with a unique mission to advance U.S. innovation. We have a very broad portfolio of research at NIST, but more importantly, a long tradition of cultivating trust in technology. And we do that by advancing measurement science and standards, measurement science and standards that makes technology more reliable, secure, private, fair, in other words, more trustworthy. And that's exactly what we have been doing in space of AI.

NIST was established in 1901 to fix the standards of weights and measures. Our predecessors created advanced standards to measure basic things such as length, mass, standards needed for electricity, light, everything that was essential for the technological innovations and competitiveness at the turn of the 20th century. And in a way, we are

following the same course working with and engaging the whole community in figuring out proper standards and measurement science for advanced technologies of our time, which I think everybody agrees artificial intelligence is in that category.

In terms of what we have been doing in this space, a little over a year ago, something like a year-and-a-half ago, we released a NIST AI Risk Management Framework or AI RMF. Directed by congressional mandate, AI RMF is a voluntary framework for managing the risk of AI in a flexible, structured and measurable way. The measurable attribute is particularly important for us, coming from a measurement science agency, because if we cannot measure it, we cannot improve it.

So, if you want to really improve the trustworthiness and responsible use of AI, we need to be able to have measure -- to know what to measure and how to measure. AI RMF was developed in close collaboration with AI community, we engage diverse groups of different background expertise and perspectives from the community that developed the

technology to the community that study the impact of the technology to running listening sessions with a community that are impacted by the technology. The framework is intended to be voluntary, rights-preserving, nonsector specific and use-case agnostic, providing flexibility to organizations of all sizes in all sectors and throughout the society, to implement the approaches in the framework. So, by design, it can be used for all of those different application that Dr. Nelson mentioned, and Dean also alluded to them.

Continuing that work in March of 2023, we released AI Resource Center as sort of a one-stop-shop of knowledge, data, tools for AI risk management. It houses AI RMF playbook that provide more sort of actionable suggestion on how to implement and operationalize AI RMF. It's cool, it's interactive, searchable, filterable. And we consider that as a work in progress as we're adding additional capabilities. For example, things such as standard hub or repository of metrics are more.

In June of 2024, again in response to the

release of the generative AI languages, we put together a generative AI public working group where more than 2,000 volunteers helped us to sort of study understand the risks that are unique to generative AI or exacerbated by generative AI. Our latest assignment, the executive order on safe, secure, and trustworthy AI builds up on all of those foundational work that we have been doing. The executive order specifically directed NIST to develop evaluations, red-teaming, safety and cybersecurity guidelines, facilitate development of consensus-based standards, and provide testing environment for evaluation of AI systems, including for dual use foundation models.

All of these guidelines and infrastructures will be voluntary resources for the use by AI community for advancing safe, secure, and trustworthy AI. I think it has been mentioned several times, I -- it's -- everybody knows that AI is the, one of the most transformative technologies of our time, one with tremendous opportunities to improve our lives, but also comes with its negative consequences and harms. That's why safeguards becomes really important.

When it comes to AI, there is a lot less we know that we should, and I think all of these conversations and what we can do is that we should try to change that. There is a lot that we can do. I'm just going to talk about five things that I jotted down last night.

So, first, we heard it in different ways that our understanding of limits and capabilities of this powerful technology is limited, so we must engage in efforts, technical and scientific efforts to advance our scientific understanding of how these models work and behave.

We heard this, this morning. But I also want to emphasize that we also must address AI's impact on people and society and planet through technical, social, and sociotechnical lenses. We should also advance research on identifying, measuring, managing, and mitigating risks, including safety, security, privacy, fairness, reliability, interpretability. One of the things AI RMF does is try to provide some sort of a taxonomy of the risks for AI systems to help with this structured, measurable approach to risk

management.

You should also, and I think this is really important, actively seek and incorporate insights from a diverse range of experts representing diverse set of backgrounds and perspectives, particularly the group that the technology is going to impact them. And data, technology does not know borders, so it's important to cultivate and strengthen international collaboration, cooperations on AI issues, but particularly on standards. Bottom line is that we want technologies that work accurately, reliably, technologies that's easy to do the right thing, difficult to do the wrong thing, and easy to recover if and when something goes off. And --

MS. FRANKLIN: If you could please wrap up your opening, so we do have time to get to the questions and answers.

MS. TABASSI: I think that's a good stop. Good place to stop. Thank you.

MS. FRANKLIN: Thank you so much.

Okay. And our final opening will come from William Usher.

MR. USHER: Good morning, distinguished members of the board and for those listening, and I'll keep my remarks brief so we can get to the question-and-answer period.

Again, my name is William Usher. I'm the Senior Director for Intelligence here at the Special Competitive Studies Project. Our mission at SCSP is to make recommendations that strengthen America's long-term competitiveness on emerging technologies as they reshape geopolitics and society over the coming decade.

Prior to joining SCSP last year, I spent 32 years as an all-source analyst and a senior executive with the Central Intelligence Agency. And I'm honored to speak with you today about the role that AI plays in the national security arena, specifically with regard to the intelligence community's mission.

As Ms. Tabassi just said, artificial intelligence stands out as a transformative force that will profoundly impact national security and global competition. President Biden's executive order last October mandated that the U.S. government departments

and agencies take care of when developing and deploying AI systems. But it also called on America to "seize the promise of this powerful new technology."

Being a leader in technology innovation is important today, but it will be vital to our nation's future economic vibrancy and to the continued resiliency of our democratic way of life in the future.

As we debate the future of AI, foreign competitors, principally the People's Republic of China, are laser-focused on taking advantage of AI for economic advantage, and to challenge U.S. leadership and the rules-based order. Beijing has openly declared its aspiration to become a leading S&T power that is able to set the pace of future scientific advancements and dictate global norms.

Now, our intelligence community has long eyed AI's potential, and they have been researching the potential uses of early forms of AI, machine learning, deep learning and natural language processing for years and have already launched limited uses of

generative AI tools.

Gen AI tools have the potential to greatly expand the scale and the efficiency with which our intelligence services can derive national security relevant insights from the growing body of digital information produced around the globe.

U.S. intelligence services, for example, will be able to leverage AI's pattern recognition capabilities to identify and alert human analysts to threats such as potential terrorist attacks, or significant military movements. This capability will make critical warnings more timely, actionable, and relevant, allowing for more effective responses to emerging threats and hidden strategic opportunities.

While the potential is great, AI also poses significant new challenges for our national security enterprise. For one thing, a host of foreign countries, including several U.S. adversaries, are already investing heavily in AI for their own national security purposes.

China, for instance, is expected to more than double its investment in AI to nearly $27 billion by

next year and $38 billion by 2027. Moreover, there is a great deal of -- while there is a great deal of attention being paid today to the creators of large expensive-to-train foundation models, the presence of several capable so-called open-source models, I'm thinking here of Meta's Llama or Mistral 7B model means that the speed at which this technology will become generally available will be very rapid. Unlike the advent of the atomic age, you will not need to be a well-resourced nation state to be able to benefit from AI technologies.

Therefore, our intelligence services must devote additional resources and effort to ascertain what foreign competitors and non-state actors are doing to develop their own indigenous AI systems, and how they intend to employ them against us and our allies. We've already seen evidence of AI being used to create believable misinformation, lifelike videos and audio files that appear authentic, that are being used to push false narratives. But these same AI tools can be used to uncover sensitive U.S. military and intelligence operations, plan more sophisticated

cyber attacks, and develop novel bio weapons.

It is this Board's mandate to provide oversight of the Federal Government's implementation of the AI executive order and this certainly poses some important new questions that need to be addressed. The forthcoming White House National Security Memorandum likely will provide the initial framing of how the government thinks these questions should be properly answered when it is released later this month.

To my mind, those questions fall into one of two broad categories. The first category is what are the parameters that will guide whether the IC can make use of any particular model. If leading-edge large language models are basically trained off the internet, which is composed mostly of U.S.-derived information, how does that affect IC agency's use of such models?

Specifically, how can agencies utilize AI and remain compliant with Intelligence Community Directive 107 concerning privacy protections? My personal view is this can be done but right now I think different

agencies are interpreting the rules differently. The second category is what will be non-acceptable uses of generative AI outputs for the U.S. intelligence community.

As we try to figure that out what are the examples of non-acceptable uses, I expect we will go through a lengthy trial and error process and formed mostly by "I'll know it when I see it" type wisdom. Some restricted areas will be obvious, such as relying solely on AI systems to target suspected terrorists for kinetic strikes, but other potential restrictions will be less obvious.

For example, imagine a scenario in which a U.S. intelligence service proposes to request that another government detain a foreign national transiting their country, which the intelligence service assesses is engaged in a terrorist plot based purely on the recommendation of an LLM AI model. What are the expectations for human review of that recommendation?

Or more challenging, what if the AI detects what it assesses to be an imminent cyberattack that

could occur any second? The AI tells you it knows exactly which U.S. computer systems to lock down to thwart the attack. There is no time to gather policymakers for a meeting in order to head off the attack. Is the AI pre-authorized to mount a defense?

We'll see what the National Security Memo says and whether a clarify thing -- clarifies things or not. But I suspect we are embarked on a long journey to determine whether and more importantly, how the IC uses AI to its advantage. I recognize the risks. But I would encourage the President, Congress, and this Board, not to prematurely tie the IC's hands because our adversaries certainly are making use of AI and we need to stay ahead. And with that, I'll close my remarks and take any questions.

MS. FRANKLIN: Thank you so much. Thank you to all of the panelists. So, we are going to try to cycle through twice, hopefully, with all board members having a chance to ask questions. And I am kicking off this round.

And I want to start with Alondra Nelson, please. So, you discussed the White House's Blueprint

for an AI Bill of Rights that I believe you lead for OSTP. And in the context of data privacy, that Blueprint lays out the need to limit data collection, to follow privacy by design principles and to incorporate robust safeguards, excuse me, robust oversight for automated systems.

And in particular, I noted that the Blueprint describes the need for heightened oversight of surveillance systems, including an assessment of potential harms, both before deployment and in an ongoing manner and to test for harm such as algorithmic discrimination. So, I'm wondering, can you provide us with any further thoughts or more detailed guidance on how in your view government agencies should conduct these pre-deployment assessments and what kind of research you've seen and what promise that holds in that space?

MS. NELSON: Thank you for that question, Chair Franklin. So, in the process of developing the Blueprint for an AI Bill of Rights, effectively what we did was distill best practices from industry, from academia and from colleagues working in government,

about these technologies. And so, you know, what we distill there is what we've learned from what people think is possible, or what they've been already using.

In the space of government and in particular, you know, obviously the pre-deployment assessment will happen in the space of acquisition and procurement. And there's quite a lot in the President's executive order, I think that, you know, that asks agencies to think about that piece of their work, as well as, we haven't mentioned yet OMB's memo on the trustworthy and safe use of AI, which, you know, suggests, as we would want to, you know, that government should be in the business, should be leading by example, and in the business of using rights-preserving technologies and technologies that if they have impact on people's safety, that we're thinking, you know, about how to do that.

I would say that the Blueprint for an AI Bill of Rights, depending on whether or not you read the PDF, or the website, has an IC carve out. And so, you know, the sort of the principles are, you know, of the -- Blueprint for an AI Bill of Rights and the

practices, as you suggest, the, you know, pre-deployment, the sort of various assessment tools that people might use, are not intended to apply to that space.

What I was trying to suggest in my remarks, as I was closing, and I apologize for running over was that, you know, as SCSP was describing in a world in which, you know, new threats can come from, you know, actors using these commercially -- widely commercially available technologies, that one of the ways that we can exercise, national security kind of prudence and oversight is actually to have oversight of commercial technologies in the civilian sphere.

And, you know, that's where I think this Board's ability to exercise oversight over the fulfillment of the executive order and other kinds of executive agency, you know, sort of mandates and levers is tremendously important.

MS. FRANKLIN: Thank you. Okay. So, I'd like to ask a question of all the panelists to hopefully quickly touch on this for us. I know, it's a big question though, which is, you know, as I noted

in kicking off this forum, we are working to scope and define our oversight of governments to use of AI for counterterrorism purposes. And I would appreciate any, you know, sort of concise thoughts you may have for us on how we should carve out an appropriate slice.

And where I'm going with that is, you know, of course, it would be completely unworkable for us to say, okay, right now we're going to start examining all the government uses of AI for counterterrorism purposes, or even to say all surveillance programs, or even to say all data analysis. So, we want to be strategic, and to think particularly about uses of AI that are more likely to present risks to privacy and civil liberties.

So, maybe I'll walk through the order, starting with Dean Souleles, if just any quick thoughts that you want to share with us on how you would advise us to carve out where we go next.

MR. SOULELES: Yeah. I think it's unworkable to dive down to any etches. I think instead what I would focus on is ensuring that the intelligence

agencies that carry out the counterterrorism measures have appropriate policies and oversight in place to manage their AI systems. So, we developed an AI maturity model for the Intelligence Community while I was still in the seat that basically told the IC how to evaluate its readiness to deploy AI systems.

And that includes things like, do you have a data -- chief data officer who's responsible for Privacy and Civil Liberties Oversight? Do you have policies and procedures in place that are such that they force you to analyze the data that you're collecting and ask all the questions that we've raised here today?

Do you understand where your algorithms come from? Do you understand the models? So, these are basic things that they must do. And I think the Oversight's Board role is to make sure that they are doing the things that they say they're doing, kind of all right, you've got this set of standards demonstrate to us that you are actually doing the things that you are saying you are doing because I don't think you're going to have the ability to get

down any of the issues, but I think if you look at it for the macro level to make sure they have the policies in place, that the policies are consistent with the President's guidance and so on, that you'll have a chance of actually doing what you're asking.

MS. FRANKLIN: Thank you. Okay. Quick thoughts from Elham Tabassi.

MS. TABASSI: Seconding and echoing everything that was mentioned. And, you know, we are a big fan of the risk-based approach. And considering the contexts of use, there is no one-size-fits-all. So, having the framework, having the question that needs to be asked, but having the flexibility to sort of set the threshold of how private is private, how bias is bias, set them based on the risk of that particular context.

MS. FRANKLIN: And quick thoughts from William Usher?

MR. USHER: Sure, Dean's got it exactly right. Right now the Intelligence Community for its human employees has clear and firm standards for their use of data and information for various purposes. And

your Board provides oversight of that activity. And I think that should be the same standards really, that are used for monitoring how they use AI.

I would hope that that does not extend to preventing the IC from taking in large language models for experimentation and examination. But certainly when it comes to utilizing the outputs from LLMs, applied to any classified data holdings, the same standards that are in place now for protecting privacy and civil liberties should apply.

MS. FRANKLIN: Thanks. Okay. So, I'm seeing that my time is up. So, I'm going to turn this over to Ed Felten.

MR. FELTEN: Thank you. And thanks to all of the panelists for your remarks and your willingness to entertain our questions. I'd like to ask all the panelists about something that has been mentioned a couple times already. And that is about the use of -- potential use of foundation models in intelligence and law enforcement. These, you know, as you know, are the largest and most sophisticated of AI models, and they're behind products like ChatGPT, they seem to

have unique capabilities that may be valuable for national security missions.

But of course, training them requires enormous investments and huge volumes of data. And this is typically done by commercial parties. But government agencies of course have detailed and strict limits on the use of information, and for good reasons relating to privacy and civil liberties. So, my question is really whether and how intelligence agencies might approach the use of foundation models, in a way that's consistent with privacy and civil liberties. Should agencies work with commercially trained models? Should they seek to build their own models? Is there some other approach? Or is this just a bridge too far from a privacy and civil liberties standpoint?

Let me start with Mr. Usher, and then go backward in order of the initial -- reverse order of the initial statements.

MR. USHER: Mr. Felten, that's a terrific question. What we have recommended is that, yes, the IC does make use of commercially available models,

because, as you noted, the cost for developing independently is going to be quite steep and I would argue probably prohibitive. I have had the pleasure of seeing one instance of the IC's use of a large language model as applies to unclassified data.

And in thinking about how it could be deployed, how they could be deployed against classified holdings, I think what we would probably expect to see as the capabilities curve kind of goes up with these models in the years ahead, the IC will kind of have to pick a version and work with it, train it to its standards for protection of civil liberties, also tradecraft standards, accuracy, et cetera. And, except that it will be a little bit behind the leading edge of what some of these foundation model developers are doing with their systems, but be more confident in their reliability, transparency, explainability, et cetera because the standards that the IC must meet for telling the truth and protecting civil liberties is and should be high.

MR. FELTEN: Thanks. Let me turn to a Dr. Tabassi.

MS. TABASSI: I completely agree, nothing more to add here, just saying that, yes, you can get the model and then try to add extra safeguards to make it up to the thresholds of the IC.

MR. FELTEN: Right. Mr. Souleles?

MR. SOULELES: Yeah, I would say that it is a great question, but it is the same question that we ask with all deployed technologies and all data analytics. It's a little more complex because, as we know, these foundation models are trained on huge volumes of data. And in the Intelligence Community, that means they're trained on data that involves U.S. persons.

I spent a good deal of time having this conversation with our attorneys and others in terms of how do we use the tools without violating the basic directives that were not allowed to perform intelligence on U.S. persons. And I would say that we need a policy that talks about that specifically and makes it clear what we can do and what we cannot do. I think we should be able to use foundational models.

As SCSP said, we should retrain them to our

own standards, but we should not eliminate the use of them, because they have been trained on U.S. person's data, because that would cut us off from a wide swath of technologies that we know our adversaries are using. And this is really similar to the conversation a few years back with NSA and telecommunications election, right, is, yeah, there's going to be U.S. person's data in there, it's incidental to the intelligence problem that we're asking. And we develop policies and procedures for using it. We didn't throw it all out completely. So, my approach would be to embrace it, put the appropriate guidelines around it and continue to use it.

MR. FELTEN: All right. Dr. Nelson?

MS. NELSON: Thank you, Dr. Felten. A few things. I mean, I raised the Clearview AI example because it suggests some of the challenges that we faced here. So, there's clearly American data in there. It's being used by Ukraine and Russia in the theater of war. And there's a lot of complexities, I think that we still need to think through in the national security space with regard to how we as

Americans want to operate in using that.

Second, I would say is that there's a -- you know, we're seeing a kind of a lot of David and Goliath, if you look at the example of what's happening in Gaza right now. So, having a big model is actually not necessarily going to be the thing that helps you win, if you like, you know, switching back over to Ukraine, when people can take commercially available drones, and, you know, that cost $1,000 and destroy a multimillion dollar tank using that, right?

So, I think that the, you know, smaller language models, the open-source models create a kind of asymmetry that we want to use the foundation models, I think, safely and effectively knowing all of the many, many caveats around them. But I think that a lot of what AI is enabling is this kind of radical asymmetry in the national security space.

And then lastly, I would say I would just point people to and commend that DARPA has just started a new program on the mathematical foundations for AI evaluations. And I think one of the first things that the IC needs to do is actually to Elham's

point, figure out how these things actually work because just air gapping the data alone or using the enterprise version, I don't think for the threshold you want for national security is actually high enough. And so I think moving forward on understanding the basic science of this is tremendously important as well.

MR. FELTEN: Great. Thank you. Thank you to all the panelists for your thoughtful answers. Let me pass the baton to my colleague, Travis LeBlanc.

MR. LeBLANC: Thank you, Ed. And also, I want to thank the panelists for joining us this morning for this important forum on artificial intelligence and how we balance it with privacy and civil liberties in the national security context.

I'd like to pick up on a conversation that Dr. Nelson was just having around Clearview AI. And I do appreciate the concerns that you elaborated on about the use of Clearview AI by Ukraine. It has been called Ukraine's secret weapon in the war. And the question I sort of have is, do you believe that the United States should refrain from using an application

or tool like Clearview AI? And if so, what do you say to those who believe that it puts the country at a disadvantage to defend itself if our adversaries are able to use these tools?

MS. NELSON: Thank you for that question, Mr. LeBlanc. We already use it widely. It's used by American police forces all over. So, it's not that's, you know, that's in some ways it's a moot question. I think it gets a little bit more complicated in the international sphere when we're talking about civil liberties and people's rights when it's American data that's being deployed in the theater of war and other spaces in Ukraine and in Russia and how do we want to think about that at a time when, you know, this administration is issuing executive orders that is constraining the flow of data for example. So, that American data should not be allowed to circulate in countries of concern for example.

So, to me the Clearview AI issue I raised because it raises a lot of fundamental questions that we don't have answers to and a lot of fundamental tensions. So, Mr. LeBlanc, I don't have any clear

answers. But I would say we're already using the technology, gets a lot more complicated when these commercial technologies are also become, you know, military technologies. And then we need to reimagine, I think the regulatory and rights regimes or either double down on them and, you know, we've got to figure that out.

MR. LeBLANC: And I guess thank you for that response. I want to follow up to ask, are there any applications of AI or any uses of AI that you believe the U.S. government should not be engaged in right now?

MS. NELSON: It depends on the context. I don't think that we should be using real-time facial recognition technology in a civilian context at all. I think that should be a red line.

MR. LeBLANC: Okay. Thank you very much. And, you know, we have a lot of professors on this webinar, but we only have one Dean. So, I want to ask Dean one question, which is, is the error rate around generative AI too high right now to be reliable? And relatedly, are there any uses of AI that should be

halted?

MR. SOULELES: So, the error rate of generative AI is pretty high. If you're talking about large language models and what we are all seeing out on the internet as chatbots, it's very easy still to do a search, do a query to the chatbot and have it return, you know, what they call hallucinations. And the important thing to understand is that despite the marketing of these tools, these are not knowledge systems, they are predictive text systems, they are trained. The idea of a large language model is it is trained on essentially the entire text of the internet. And it's able to produce in a remarkable way an English-readable sentence and paragraph and sentences based on all the text that has already been produced. But it doesn't think in any of the ways that humans think. So, we should be very careful when we use those sorts of things.

I would say though that intelligence analysts are already used to working in a probabilistic world. They have to have data that is checked against other data. They never take one source of data as the

ground truth. So, I would say I wouldn't prohibit the use of them. But we need to understand how they're being used and we need to not use them as a source of ground truth. Just like I wouldn't use Wikipedia as my ground truth for data. It's a, you know, crowd sourced encyclopedia.

So, are there areas where I would say we should not use it? I mean, when you get down to decisions of targeting individuals for lethal action, then we cannot see that to the automation today, and probably not ever, right? Bob Work talks about the need to have AI implement commander's intent in warfare, just as we implement commander's intent when we issue orders to troops. And the same rules of the road should apply.

But at some point, the gap between the commander and the execution of that command is broken. And you have to rely on the thing that's executing the command to do its thing to make sure that it's built that way. So, hope that's helpful.

MR. LeBLANC: That is indeed helpful. And, you know, your reference to targeting and, you know,

prohibitions on targeting individuals for lethal action reminds me for one last question for Dr. Nelson, which is, you did say in your opening statement that we had to talk about drones. And I just want to give you a moment in case you'd like to discuss any of the concerns or other issues that you'd like to cover around drones.

MS. NELSON: Thank you for that, Mr. Leblanc. So, I think drones is also another case study for thinking about the challenges we face at the intersection of the national security and civil liberties piece. You know, as I said, these are relatively inexpensive technologies. As we're seeing in Ukraine, they're being kind of refitted with, you know, cameras and other things to be used for -- to be made sort of as semi-autonomous weapons, you know, with that are partly guided.

And so, then we have growing capabilities out to swarm drones to have them act as both as agents and collectively. And so, then that runs into questions that we have around, you know, lethal autonomous weapons, right, like conventions that are existing

around that, you know, regulations, and the international relations space that exist around that and what we might need to think about and new ways about that.

And, you know, Ukraine is an interesting example because it's already been, and I'm sure will continue to be, a really important technology capital. There's great technologists there. And so, part of what we're seeing is about the capabilities of this particular community to be able to take drones and make them into warfare. But I guess the challenge that this board faces, I think, is that back and forth, those back and forth vectors between civilian and military technologies that make these questions open questions, rather than I think clear-cut answers.

MR. LeBLANC: Thank you. And do you think that there would be a good use of the Board to look into how DOD uses artificial intelligence in lethal strikes?

MS. NELSON: Yes. I mean, I think others, either Dean or SCSP had -- has already mentioned this. I mean, the DOD has been quite a leader. And when I

first came to OSTP, as a day one person in the Biden-Harris Administration, I believe DOD was one of the few agencies that had already released a set of principles and guidelines around AI. So, like very forward-leaning here. And the question then becomes, you know, how do we ensure that people are actually doing that, which colleagues have already mentioned? And I think that's a perfect place for this particular board to exercise its oversight.

MR. LeBLANC: Thank you. Beth Williams, I'm passing the baton to you.

MS. WILLIAMS: Okay. Good morning. Thank you, Travis. And thank you to all of our panelists. Really appreciate you being with us here today for the forum.

So, my first question is actually to Mr. Usher. So, one question is how can the intelligence community leverage commercially available and open-source resources and still protect classified information that's used for developing, deploying, and using its own in-house AI systems? One of the things, you know, in the reading was the concern that AI

systems that are in use by the IC could be reverse-engineered to divulge classified sources. And that's obviously a big concern, not only for national security, but for the privacy and civil liberty and safety of the sources themselves. So, could you talk a little bit about that?

MR. USHER: Absolutely. And this is a terrific topic for exploration because as these systems become more and more capable in the years ahead, I predict that they will be viewed eventually as critical national security assets. Some have made the analogy to the Manhattan Project. They will be so valuable that we'll have to, you know, bury them in a deep vault and protect them with several rings of security.

Mechanically, today, the way that that is done is by putting them on secure servers, which have built-in protections. It's how we onboard whatever piece of software or data that we wish to use in the Intelligence Community and keep it protected. And they're pretty well-established security protocols. Access to those systems are limited to people with a

security clearance, et cetera, et cetera.

You mentioned how adversaries will be viewing them. And I certainly think they will be seen as targets, probably targets by adversaries' own AIs. So, one thing to think about is sort of an AI versus AI intelligence for where their AIs are seeking to gain access to our secure servers to pollute, debilitate or otherwise wreck our AI systems. This is a growing area of research known as adversarial AI. And there are various techniques that one could use to attack another's AI systems.

You could mess with the training. You could mess with the data. You could give it instructions in the algorithm to generate false or misleading outputs. And there are -- any number of techniques and the entire AI stack that we would deploy for intelligence purposes will need to be protected. And that will include physical protections, protections for the personnel who have access to it and certainly protections for the algorithm and the data.

MS. WILLIAMS: Thank you very much.

My next question is for Mr. Souleles. One

question that our Board looks at when we're doing oversight projects or looking at systems is, what is the value? So, are there systems -- are these systems producing value that justifies their use? And you in your opening statement, you mentioned that AI facial recognition and other biometrics are becoming increasingly useful for keeping track of known terrorists who are trying to obfuscate their persons. And so, I'm interested in that in the context of biometrics in aviation.

How do you look at that system? Right now the system is not comparing faces to any terrorist database. Do you think that that is a concern for the overall usefulness of the system?

MR. SOULELES: I do. I mean, I think that we should look at where these systems work and where they don't work. To Alondra's point, to Ms. Nelson's point, the -- there are limitations based on the way these systems are trained. But the important thing with all of the technologies that we deploy is that we understand the four corners of the box in which they work and don't try and use them in the off label

methods for which they were used.

The early issues with law enforcement using facial recognition were, in my view, similar to off label use of medication. They took something that was built for one purpose, they didn't understand the limitations and they immediately deployed in another purpose and it didn't work. And we should not do that. We should understand how we use those sorts of things. But facial recognition and biometric recognition are some of the most important tools that we have for identifying and keeping track of bad actors, frankly. And that's, you know, and that -- and we need to continue to explore, but we need to do it in a way where we are always asking the question about, you know, where it works and where it doesn't work and what the risks are.

It really is a different domain from law enforcement. And we need to keep in mind that the rules of engagement are different. And it is not a civilian use of the technology. It's a use for keeping the nation safe from the worst actors on the planet that want to do us harm.

MS. WILLIAMS: Thank you. My final question is for Ms. Tabassi. You know, one of the concerns is that the AI wont be accurate, right? And you see that kind of at a very basic non AI level with credit reporting. One reason that people are now allowed to request their credit reports and look at them is to -- so that they can look for inaccuracies so that -- to ensure that the data is correct for which decisions are being made.

From a NIST perspective, do you have any recommendations for ways to increase the accuracy? Are there ways in a national security context that would allow people to confirm that their information is correct if it's being used?

MS. TABASSI: Right. Thank you so very much for that question. You're pointing out to the important topic of evaluations and being able to measure. First, we need to know what that accuracy is. And for a lot of these systems, we don't know. We have the anecdotes and experiences that they hallucinate it and give the right answers and so many other risks, but we don't quite know how to measure

accuracy, the false positive, false negatives? Are they still applicable or not? Or do we need a different metrics?

And also, measuring just for the accuracy in the laboratory setting doesn't gives us a good understanding and wholesome understanding of all the risks, harms and impacts that can happen in the actual context of use. So, from this NIST point of view, we emphasis a lot on the measurement and we -- not only measurement in the laboratory setting, but also measurement in the actual native real world context of the use of the algorithm.

We also know that all of this -- the science of AI evaluations is at a nascent stage. And that's where we need to put a lot more research and understanding on how to do this. Hope that was helpful.

MS. WILLIAMS: Thank you. And with that, I'll turn it back to Sharon.

MS. FRANKLIN: Thank you. Okay. So, we're going to try and have a more lightning round this time working through each board member getting a chance, I

think for just one question. So, my question is to both Alondra Nelson and Elham Tabassi, having worked on the OSTP and the NIST frameworks in this space. And multiple people have spoken already about how, you know, national security, of course, raises unique considerations.

And Dr. Nelson, you mentioned a carve out even in the blueprint that you worked on for national security. But I'm wondering if you can point toward with -- when you do think about safeguards that can be put in place to protect privacy and civil liberties in addition to, of course, the basic safeguard of having robust oversight. Are there any particular types of safeguards that you would point to beyond oversight in general that you think are or can be particularly effective in the national security space understanding the particular considerations that are in involved in that context that are different from other uses of AI? So, maybe we can start with Dr. Nelson, and then move on to Elham Tabassi.

MS. NELSON: Yeah. Just briefly, I think it Senator Rounds, you know, one, they're kind of talking

about metrics. I mean, are the tools, the use of the tools doing, you know, fulfilling the mission or not and how do we, you know, collect that data and analyze that data? I mean, that remains the, I think the key way to answer that question. Obviously internal to the IC, we're getting better and more robust tools at doing auditing of systems both, you know, before you deploy them and after.

And I think part of what's been encouraging about the last couple of years in the space of AI governance and evaluation more generally is that we're starting to see an ecosystem of different kinds of auditing, red-teaming, both adversarial and otherwise kinds of tools that allow us to know a bit more about the two -- about how the systems work.

MS. TABASSI: Thank you for the question. Very quickly. So, it's important to test. My first answer is test, test, test at all of these stages of the lifecycle. But also, it's important to get a lot of these considerations into the design of the system. So, instead of just wait until later and then test the system to see if it's private enough or not, what are

the things that -- what are the mechanisms and techniques that can be implemented and designed into the tools, the technology to make it, for example, all of the work around the privacy enhancing technologies to build the technologies, their models that are inherently more secure, more private?

So, both at the time of the design and development and do more testing across the whole AI stack or lifestyle. Thank you.

MS. FRANKLIN: Thank you. Over to Ed Felten.

MR. FELTEN: Thanks. I have a question for the other two panelists. Mr. Souleles and Mr. Usher, based on your experience working in intelligence agencies, the conversation about AI privacy and civil liberties is often framed as a kind of reactive story that AI comes along, it erodes privacy and civil liberties and we look for policy interventions to minimize the damage.

Well, my question is about how we might flip that script. Are there proactive ways to use AI within government to strengthen privacy and civil liberties and to reduce other kinds of risks? And I'd

like to ask the question both in general and also specifically, what should government agencies be doing toward that goal? Mr. Souleles first, please.

MR. SOULELES: Yeah. I think I would start by, again, going back to definitions, what do we mean by bias and systems, right? Often, when we have a conversation with privacy and civil liberties folks, bias means we're denying or causing some harm to some class of individuals based on the fact that the data itself is biased.

Computer systems don't have that concept. And data analytic systems don't have that concept. All data analytic systems are biased. We make biased decisions when we choose what data to include and what we choose what data not to include. And we make biased decisions when we decide what questions to ask or what not to ask. So, when we ask, can we eliminate bias in our systems? It's actually a false question. You cannot because all data analytic systems are biased. There's neither good bias nor bad bias. There just is bias in the systems.

The important thing is to understand the

biases and deploy them where they -- and only deploy the systems where we know they are workable. And that's complicated and not always obvious. So, I would say that what the community can do is to begin to develop more and more and better data analytic tools to describe the biases in the data that it already has and to make sure that we set appropriate guidelines around the use of that data.

MR. FELTEN: Mr. Usher?

MR. USHER: Sure. I'll just build on what Dean was saying and actually allude to a point that the Dean made earlier in his remarks that, you know, the machine learning systems applied against a rule can be fairly effective. So, you asked to flip the script. The IC right now has a set of guidelines and rules that it follows with regard to the use of U.S. person's data. And humans operating today in the Intelligence Community sometimes make mistakes.

And they put wittingly or unintentionally such data into a report or an assessment or something like that. And it takes other humans to catch the error and to remove that information. One could

imagine that an AI-enabled tool would be much more efficient and fast and would have perhaps a greater scope of reach across everything that the, say the National Security Agency is producing to make sure that it complies with established guidelines with regard to privacy protections or other guidelines with regard to quality, transparency, application of tradecraft and proper classification.

MR. FELTEN: Thanks. Onto Travis LeBlanc.

MR. LeBLANC: Thank you, Ed. I have a follow-up question to Mr. Souleles. I completely agree with you that when it comes to data sets and the use of AI that bias is likely to be -- bias is inherent, not likely to be inherited. It will always exist. Where I do digress from your view is I do believe that there is bad bias that is out there. And whether you agree with that or not, it's apparent that someone is deciding what bias is acceptable when deploying an AI system.

How can we ensure that that decision-making is more transparent even in the national security context so that at least the public or other decision

makers can know that a particular calculus or acceptance of bias was being made? And you're on mute.

MR. SOULELES: Sorry. I think it's important that we require that our deployment of systems that our data-based and are trained on data that is collected from any source, that we have pretty strict guidelines on how we analyze that data and classify it and determine that and that we have policies and procedures in place to actually assess the biases in the data.

For example, and by the way, I don't disagree with you. I agree that there is bad bias in our data. If you were to train a system to make loan decisions based on loans that were made in the 1950s, in the south, you would get a very biased system and you would reinforce that bias.

That is not the kind of bias that any of us want to see reinforced. That's why I say it's really, really important for us to understand the data that is being put into the system. And there's no magic bullet here. It requires people with data science and

analytic skills. It requires social science skills. It requires a whole broad range of skill so that we even -- sometimes we don't even know the question to ask. And if we don't know the question to ask, we're never going to know what the -- and we may not know it until it produces a bad result. And that is problematic.

But as SCSP mentioned earlier, I don't think it's a reason to stop. I think it is a reason to continue to ask the question and ensure that our agencies are actually doing the things that they say they are doing. You know, I sent out a summary of the U.S. intelligence community's privacy and some of these (phonetic) guidelines, and they do -- they say everything you want them to say, right? The question is, are they actually doing it? And do we have enough oversight to make sure that they're doing the things that they say they are doing?

MR. LeBLANC: Thank you very much. And I will pass it on to Beth Williams.

MS. WILLIAMS: Thank you very much. So, for a final question, you know, I thought I would turn it

all over to you to get your thoughts on this. When we discuss AI and I think in the general discussion it's often very esoteric, it gets to very high level principles and many people don't understand what it actually means to use AI in a national security context.

And so, I'm wondering if you can share your view of perhaps the most promising use of AI in a counterterrorism situation. If all of you -- if you have examples that you could share, that would maybe put some meat or explanation to what this actually is. And we can start with Dr. Nelson.

MS. NELSON: So, I would go back to my DART mission example. I mean, it's intended to be about an asteroid. But you could imagine that technology, that's mean that shifts the trajectory of something that's coming towards, you know, the United States or the planet could be used for, you know, weapons and these sorts of, you know, kind of spatial warfare. So, I think that's -- I'm a big fan of that one.

MS. WILLIAMS: Okay. And Ms. Tabassi, so, our board, actually it's supposed to be looking at

counterterrorism applications and we can look at programs that also touch on counterterrorism. But do you have any ideas with regard to how it can be used specifically for counterterrorism?

MS. TABASSI: I think everything that AI is good for and that is trying to understand and analyze a lot of data and make -- improve the data analytics. I cannot think of a particular example, but anything I have found often of what Alondra just said about the (inaudible).

MS. WILLIAMS: Okay. And Mr. Souleles?

MR. SOULELES: Let's see here. Yeah. So, imagine that you are a young analyst working in the National Counterterrorism Center at the Office of Director of National Intelligence and your job is to come in, in the morning and read your (inaudible) of all of the reporting that's happened overnight. And there may be many thousands of reports that have flown in from all around the world, both open source and classified.

And your job as the first order, first guideline is to sort through all that and find out

which of those things might be important to the question of the day. And the question of the day is a different question today than it was yesterday, right? The difference between September 10, 2001 and September 11, 2001, you're asked to analyze a new question today.

And your job -- and you're the most junior analyst in the department and your job is to be on that watch and just pick out the things -- you know, we call it swipe left and swipe right, you know, for reporting. And your job is just to pick the things that are most useful for the next level up to actually read and recognize. And you have maybe two or three seconds to look at each report before you make that decision.

That's an area where computer analysis, summarization, all the kinds of things that these things we know do actually really, really well and they're not making any assessments or judgments, they're just saying, let's create a sieve so that the human analyst gets to look at the most important things and not the least important things." So, I

think that's a very specific recommendation for the kind of things that we could use today and would be actually of great benefit.

MS. WILLIAMS: Thank you. Mr. Usher.

MR. USHER: I'll give you a real-world example and we don't have to go too far back in history to find it. But in 2018, Israel's intelligence service, the Mossad, according to press reports, sent a team into Iran and secretly raided a vault that contained the nuclear archive for Iran's nuclear program. The team sat on the ground for several hours and stole about 20 percent of that archive.

And according to the press accounts, that included about 55,000 documents and about 55,000 CD-ROMs with audio and video files, almost all of which was in Farsi. And they brought that back to Israel for exploitation. And you can imagine the pressure that was on the Mossad analysts who were charged with taking that raw data and trying to make sense of it to answer the urgent question at the time as to whether or not Iran's nuclear program, which had existed from

1999 to 2003, was in fact continuing, or perhaps the world misunderstood where they left off in their capabilities.

The Israeli team charged with making sense of that vast amount of data took months to process that information. With today's AI capabilities, and I'm talking here broadly about even earlier versions of AI such as machine translation, which is not quite effective, the first two steps in that analytic process, translating the material into Hebrew or other languages, and identifying salient points within that data that the analysts should look at and in which priority can happen now within minutes, if not seconds, right?

This is a tremendous advantage when dealing with the intelligence challenges of the future, where we'll be looking at large datasets, entire computer networks, or a foreign country's AI stack, where it's impossible for humans, even large teams of humans, to go through that accurately, reliably, quickly. AI is a real boon to the intelligence community in a situation like that.

MS. FRANKLIN: Thank you very much. Okay. So, that is going to bring our first panel to a close. I'd like to thank each of our panelists for this first panel, for sharing your insights with us. And for our audience, we are now going to take a short 5-minute break and we will then return for our second panel. Thank you

(Recess)

MS. FRANKLIN: Okay. Thank you. We are now back for our second panel. And I'd like to welcome them all here. We will again, with this panel, proceed through the panelists in alphabetical order for brief opening statements, and then move on to board member questions. This time the board members will reverse the order of board member questioning.

So, our panelists for this panel are Miranda Bogen, who is director of the AI Governance Lab at the Center for Democracy & Technology; Clare Garvie, who is counsel at the National Association of Criminal Defense Lawyers; Jamil Jaffer, who is director of the National Security Institute at George Mason Law School; and Peter Winn who is acting chief privacy and

civil liberties officer at the Department of Justice.

So, first to Miranda Bogen, for your opening remarks.

MS. BOGEN: Thank you so much. And thank you to the Privacy and Civil Liberties Oversight Board for the opportunity to provide comments today about the privacy and civil liberties implications of AI. My name is Miranda Bogen, as mentioned, and I'm the director of the AI Governance Lab at the Center for Democracy & Technology, which is a nonprofit and nonpartisan organization that defends civil rights and civil liberties and democratic values in the digital age. The AI Governance Lab works to develop actionable and practical efforts to govern AI -- the use and development of AI responsibly.

Prior to joining CDT, I worked with developers and deployers of advanced AI systems and machine learning models at Meta, where I was directly involved in defining processes for managing risks presented by these technologies, and building approaches and guidance to encourage the adoption of more responsible AI development practices.

The newest AI powered methods and tools can offer benefits for organizations and government actors. But we urge caution, especially when considering uses in high stakes contexts such as national security and counterterrorism, given the many well-known, but unresolved risks that AI systems pose to people's rights and safety.

First, intelligence agencies may seek to use AI to help analyze and act on huge swathes of text, audio, image, and video intelligence. We're deeply concerned, however, that without appropriate safeguards and oversight, this technology will be deployed to facilitate and dramatically expand indiscriminate surveillance and increased reliance on automated tools to inform national security activities. Incomplete, unrepresentative, and biased training data can lead to erroneous discriminatory and harmful outcomes, and even functional AI tools can lead to the suppression of dissent and the oppression of marginalized groups.

In addition to embedding pernicious biases that may be challenging to detect, in many cases, AI

outputs are highly arbitrary because the process of training machine learning and AI models unavoidably involves a significant amount of randomness, which risks leading to erroneous outcomes that will disadvantage and harm people.

Second, AI powered systems remain inherently unreliable and difficult to scrutinize, making oversight critically important. Simply put, the intelligence community should not assume that AI augmented analysis is by default more accurate than human analysis. AI systems remain vulnerable to subjective judgments reflected in training data, as well as to the human interpretation of outputs, hallucinations, and changes the system settings that can lead to increased errors and flawed outcomes.

To maintain some degree of confidence in the performance of an AI system, independent oversight should involve making sure that agency's focus on training data used to develop AI systems was lawfully and ethically gathered, and is relevant to the system's intended uses. Supporting transparency into how systems are customized, fine-tuned, and validated

for national security purposes, and maintaining visibility into how these systems are integrated into operational work and how their outputs are acted on in order to prevent the erosion of safeguards against errors and biases.

Third, ensuring human decision makers with subject matter and domain expertise can and do maintain meaningful oversight over the use of AI systems, will require proactive effort. National security institutions must put in place internal as well as independent governance mechanisms to promote the responsible use of AI. They should clearly assign decision making and internal oversight responsibilities, require review and approval by high level officials for the procurement of systems and scrutiny of use cases that present particularly high risk. Privacy, civil liberties, and legal officials should be given comprehensive visibility into how departments and agencies are using AI and must be included as part of the decision-making process through the AI development, procurement, and deployment lifecycle.

Fourth, AI should not circumvent rules and safeguards established for intelligence agencies and personnel. For example, if Congress requires court approval before the results of U.S. person queries, a 702 collected communications can be reviewed. Intelligence personnel might seek to use AI to circumvent such a rule by tasking an AI system to review the communication based on the position that no human review was conducted and thus no court approval was required. Things like this should not be permitted.

Finally, PCLOB should assess compliance with insufficiency of existing executive policies on agency's use of AI. As an independent oversight agency with access to classified programs, you are uniquely poised to assess the effectiveness of administration policy on agency's use of AI, including ensuring that the forthcoming memorandum on national security uses of AI is applied narrowly, only to those uses of AI exclusively centered on national security. Other AI applications are subject to the OMB governance memorandum.

As intelligence and national security agencies deepen their pursuit and investment in technologies like artificial intelligence, the careful consideration of privacy and civil liberties implications of AI systems is both necessary and urgent. Independent oversight and expertise will play a critical role in ensuring that decisions around the appropriate use of AI power tools remain grounded in human rights and core democratic values.

Thank you.

MS. FRANKLIN: Thank you. Next, Clare Garvie.

MS. GARVIE: Thank you so much for inviting me to speak on this panel with you today.

I want to start with an example because I think it's a helpful illustration. So, 10 years ago, an Israeli company called Faception began marketing an AI based system to identify possible future terrorists in real time, without any prior intelligence of the person required. The tool, according to the startup, could predict someone's propensity to be involved in future acts of violence, based on an analysis of their

facial features, captured in video at a distance.

When asked by a Wall Street Journal reporter back in 2018 about the foundational validity or reliability underpinning the tool, Shai Gilboa, co-founder and CEO of the Faception stated, "I need to emphasize that there is no scientific evidence for the terrorist classifier." Nevertheless, this system continues to be promoted, and is used by at least two, as of yet, unnamed country's defense agencies. The company also markets tools to identify possible white-collar criminals, pedophiles, brand promoters, bingo players, and academic researchers.

I highlight this tool not because I suspect that U.S. is one of the countries using it. I have no evidence one way or the other. But because I think it illustrates many of the privacy, civil liberties, reliability, transparency, and other concerns with AI that we're here to discuss today. And we've already heard a fair amount about including; one, the often-unquestioned impulse to see AI as providing a solution to all intelligence, national security, or law enforcement challenges. This ability to identify the

next potential plot, screen travelers, gather evidence, without necessarily considering the true costs or evaluating viable alternatives.

Two, closely related, the fact that AI may over-promise and under-deliver, put simply, we risk deploying junk science in an extremely high consequence environment, both on the national security side and for the people investigated or denied access or benefits based on AI determinations.

Three, the threat of entrenching existing and often biased heuristics about who or what constitutes a threat. Faception's terrorist classifier appears to look for Middle Eastern male faces. It failed to flag Ted Kaczynski as a possible threat for example, and at least initially, it was not trained on women at all. This bias is well documented across facial recognition deployments, but is in no way unique to facial recognition systems alone.

Four, the increased reliance on AI to define and identify what constitutes anomalous and often suspicious or probable cause level behaviors or people risking supplanting human and judicial determinations

of probable cause, and in some cases, even guilt.

And five, exacerbating the others, a tendency for AI systems to add layers of opacity onto already deeply non-transparent sectors, like intelligence and national security.

To narrow this focus somewhat, as this board is of course acutely aware, two of the core mechanisms to ensure privacy and civil liberties in the intelligence and national security space are; one, the minimization of collection, retention, and dissemination of U.S. persons' data. And two, transparency and oversight.

In evaluating national security applications of artificial intelligence, I urge the board to consider that AI and the promise that many of its applications hold out, is in tension with these mechanisms. Many AI systems brought the ability to ingest and make sense of vast quantities of disparate information about people, associations, behaviors, and more. This combined with system needs for large representative training datasets, creates an incentive for more, not less data collection, retention, and

dissemination.

On the transparency and oversight mechanism, the black-box nature of algorithms coupled with trade secret claims that accompany private sector development of algorithms often leave agency users themselves, not to mention the public, uninformed about potential sources of error and bias and threats to privacy and civil liberties. This is exacerbated by the rapidly evolving nature of AI based systems, a pace that I believe our current structure of privacy impact assessments, systems of records notices, and other transparency mechanisms have little hope of keeping up with.

I further suggest first and foremost, orienting to the question of whether is a tool necessary? And if it is, is it necessary that that given tool be AI based at all? Or does the data collection transparency, reliability, and bias concerns posed by the system and introduced by the AI component outweigh the purported benefits? I also encourage the board to push executive agencies to think critically about whether the current oversight

and transparency structure is adequately responsive to the realities of AI, its pace of development and deployment in the face of those harms.

Thank you so much. I look forward to answering your questions.

MS. FRANKLIN: Thank you. We'll next hear from Jamil Jaffer.

MR. JAFFER: Thank you, Chair Franklin, and board members for having me here today. My name is Jamil Jaffer, I'm the founder and executive director of the National Security Institute at George Mason University's Antonin Scalia Law School. I'm thrilled to be here today at this forum as PCLOB takes advantage of its statutory responsibility to take action, analyze reactions of executive branch that are focused on protecting the nation from terrorism.

Today, the threat of terrorism is extreme. We just heard in the last few weeks from the FBI director that he believes that he is hard pressed to think of a time at which so many different threats to our public safety and national security were so elevated all at once. We know the world is on fire.

We see the wars in Ukraine, the war in the Middle East, a potential threat from China in the Indo-Pacific. And the FBI director is telling us that the threat from foreign terrorists has risen to a whole another level since the October 7, 2023, terrorist attacks on Israel by Hamas.

Director Wray went on to note that there's already a heightened risk of violence in the United States before October 7. And since then, the FBI has seen a rose gallery of foreign terrorist organizations call for attacks on Americans and their allies, raising concerns. Not only that individuals and small groups will draw twisted inspiration from what's happened in the Middle East, but there's increasing concern by the potential for a coordinated attack here in the homeland. A (Inaudible) attack conducted in Moscow by ISIS K, ISIS Khorasan, that took the lives of over 150 or nearly 150 and injured over 500. The threat is extreme.

In fact, Graham Allison and former deputy CIA director, Michael Morell, reported in foreign affairs just last month, that the terrorism warning lights are

blinking red. The United States faces a serious threat of terrorism in the months ahead. This is an extreme situation. This is not a time to step cautiously and pause on our questions about whether we should take advantage of the AI revolution to counter terrorist threats. Today is a time where we must lean forward. Now, we must do so in the context of our values and the protection and privacy -- and protection of the privacy and civil liberties of Americans. That is critical.

But the way to do that is to not slow down what we implement, to not think hard about (inaudible), but to lean forward and to think about how we can build AI capabilities for the national security community, for the counterterrorism community in a way that bakes trust, safety, and security in from the jump at development, in deployment, and on a going forward basis. That doesn't require going slow, but it does require thinking hard about trust, safety, and security.

So, how do we do that? How do we bake in trust, safety, and security right from the jump?

Well, we're not writing on a blank slate. Luckily, we have a long-time scenario of dealing with these questions in other domains; cybersecurity, counterterrorism, and other domains where we bake trust, safety, and security and at the outset, we need to do more, we need to get better. But the government's already doing this. DHS has secure by design principles, resilience by design principles for software. NIST, as you heard earlier today, has reliable AI standards. NIST has reliable cybersecurity standards. A lot of these standards are built on what industry is doing already and how industry might lean forward.

The government can incentivize the adoption of safety, trust, and security in their systems by using their buying power. The government can provide incentives in the form of tax relief, they write incentives in the form of liability and regulatory relief. The government can provide incentives in the form of grants to companies and organizations that are building these capabilities to make them more trusted, to make them more safe, and to make them more secure.

And in fact, investors and innovators have an incentive for baking trust, safety, and security into their systems. It makes the products that they build more likely to be adopted by the government. And by industry if they're trusted, safe, and secure. This idea that we need to treat AI, like it's a global pandemic or like it's a nuclear weapon, as some have suggested, is simply wrongheaded. AI has the power to be transformative, we ought to take advantage of it, particularly at this heightened threat level.

And just to demonstrate that, in fact, investors and innovators have the incentive to invest in this, the venture capital firm that I work with, Paladin Capital, led a group recently of a dozen venture capital investors, along with the NATO Innovation Fund, signing a series of principles around investment in trust, safety, and security. There's a growing market in this space. This is not a time to go slow. It's a time to lean forward, but to do so in a way consistent with our values and the protection of the privacy and civil liberties of Americans.

Thank you for your time, and I look forward

to your questions.

MS. FRANKLIN: Thank you. And now we'll hear from Peter Winn.

MR. WINN: Thank you, Chair Franklin. And thank you to the other members of the board. I look forward to your questions.

Before I begin, I just wanted to say something that I think I've drawn from comments that some of the other thoughtful commenters have made, which is AI is a tool and it's used by humans. We have a lot of laws out there that apply to humans. But it's not as if those laws cease to apply when you're using AI. Those laws still apply. If a law forbids discrimination in certain ways, the use of AI to discriminate will violate that law. If an AI -- I mean, there's a lot of examples. My favorite might be the recent example of an AI program that used an actress' voice. Well, AI didn't make the rules about inappropriate appropriation of a person's identity, without their permission, go away just because you're using an AI program.

So, what I'm getting at is that the

Department of Justice, if we, you know, we collect a lot of information, and we have to use new technologies in order to keep the public safe and protect national security. If we lose the trust of the public when we're doing that, we're going to lose the authorities that we depend on to collect that information that we need to protect people. So, trust is mission critical. And the best way to lose trust is not to comply with the laws that apply to us or not comply with the frameworks that we've adopted.

Now, I'd like to spend most of my time talking about the recent executive order on AI. There's been a mention of the national security memorandum on AI that's part of that executive order. I'm not in a position to discuss that because that's still being deliberated. But I would encourage the board to refer back to the 2020 AI framework for the intelligence community that was developed, I guess, now nearly 4 years ago, and how thoughtful and forward leaning that framework is, and how so many of the frameworks concepts that were developed at that time have been now, even see them in the executive order on

the safe, secure, and trustworthy development and use of AI.

The guidelines and practices aligned with the NIST AI risk management framework are extraordinarily helpful tools in the development of AI. The efforts to mitigate the risks of inappropriate algorithmic discrimination that can may be exacerbated by AI. As some of the commenters pointed out, you can't eliminate bias, you're trying to mitigate bias that you don't want to have happened, inappropriate bias or unwanted bias. All systems are going to be biased.

The OMB directives that have been issued implementing the executive order are extremely helpful. So, the Department of Justice has so far designated Jonathan Mayer as our chief AI officer. We've launched the Emerging Technologies Board. And we've complied with our AI use inventory. And it's up on the department's Open Data website. We're looking to include AI assessments as part of the system of procurement and development process. And we encourage the board to review the NIST AI risk management framework and playbook. They're not prescriptive, but

those NIST tools, we have found, represent a really excellent roadmap for any organization wishing to engage in conscientious implementation of this new technology.

In April, NIST released a draft publication to help manage the risk of generative AI. And the generative AI profile can help organizations identify unique risks posed by generative AI and to mitigate those risks in a way that aligns with that organization's goals and priorities. That profile identifies a group of 12 risks relating to generative AI. Three of those, I think are key, having to do with data privacy, information security, and general information governance. The NIST framework provides a set of actions to help organizations identify, measure, map, and manage those risks consistent with that risk management framework.

AI is a novel, emerging technology, but its use cases are generally understandable. And the existing technology neutral legal structures, the government is already subject to, are excellent ways in which we need to implement the AI just as we've

navigated other prior technological advancements.

For example, the department is required under the E-Government Act of 2002, to conduct privacy impact assessments, whenever it implements a new information technology. Now, the last time I heard, AI was an information technology. So, we're going to be applying our existing sound privacy impact assessment frameworks to the requirements in the EEO to addressing the unique risk factors of AI in a rational and responsible manner.

Whenever we implement AI systems, such as facial recognition technologies, we always require a human to be in the loop, where the AI is used to make determinations about individuals. This is a longstanding standard within the department's policy development and practice. And we fully intend to be implementing that basic requirement of having humans in the loop when we're implementing AI programs.

I think Dean had a really insightful observation that the usefulness of AI is a relationship between machines and humans. And it's governed by rules. When Deep Blue defeated Garry

Kasparov, that was not a defeat that took place because the machine was better than the human or the human running the machine was better than the human. But because the interface and the rules governing that interface, that was what made the difference, that made the difference that provided the human machine interface such a powerful tool that it defeated the world's grandmaster. And I think keeping focused on the human relationship to the AI programs that we're going to be implementing is the key to advancement of, you know, those technology.

Thank you.

MS. FRANKLIN: Thank you. So, we'll now start with questions by Board Member Beth Williams.

MS. WILLIAMS: Okay. Thank you very much. And thank you to all of our panelists for being here today. We really appreciate your views and your expertise on these questions.

So, my first question is actually for Mr. Jaffer. You talked about the trust, safety, and security. And, you know, focusing specifically on trust, I always think that one of the issues with AI

is kind of a confidence problem, right? The only thing worse than somebody having the wrong answer is the guy who's also very confident that his answer is right. And I think that could be a problem with AI in that we, you know, if people believe it to be 99.999 percent accurate all the time, they're putting trust in the answers that it's giving them or the outputs that it's giving them.

So, my question to you is, are there ways that you've thought about that we could address that confidence issue? Is there a way to put a maybe next to your answer a confidence estimate? Or are there like other AI programs that should be labeled, layered on top of existing AI programs to give human users analyses of how likely to be correct certain outputs are?

MR. JAFFER: Yeah, it's a great question, Member Williams. You know, the -- I think part of the challenge with when you talk about AI and its capabilities is, we heard earlier about sort of idea that we're sort of associating human values with AI because it sounds and feels colloquial. So, we trust

it the way we trust a human, I mean, trust that it's not sort of freelancing, but even humans, you know, make things up, right?

In a lot of ways, the way that AI works by associating words with other words that we -- that it puts together may very well be how we interpret things in our brain as well. We're not actually sure when a person tells them, whether they're actually telling us the truth or not. But we judge them based on a lot of other factors. There's got to be ways to do the same with AI.

We're not going to get to a point, I don't think, where we're going to eliminate all of the "Hallucination problem." What we can do, however, is create capabilities like you say, that provide confidence assessments that allow AI models to ingest the data from other models and regurgitate what they see is the right answer amongst a variety of them.

Sometimes with some AI models, if you look at Google Gemini, you'll see it'll give you three different versions of the same answer to see which one you are more confident in. And if you had confidence

metrics associated with that, that might actually give you more to pick from.

And in addition, there are now capabilities, some of which venture capital firms like ours are investing in that actually look at AI output and say, are we getting the right thing? Is the model working in the right ways, where you can sort of put your model in and ensure your model is doing what you want it to do? Those aren't going to ever be a 100 percent, but you can get better and better over time and that's a way of creating confidence also.

At the end of the day, though, I think people have to recognize that this is a tool and the capability, it's not an answer. So, you know, you just had, you know, Peter talk about a human in the loop. We heard about that earlier as well. There's also this notion of a human on the loop, which is to say, there are some automated decisions to be made, but a human can intervene and stop a decision or walk it back if need be.

So, there's a variety of ways that we have of humans engaging with AI. But more often than not,

what it really is it's not a substitute for human judgment. It can't and shouldn't be. It's simply a supplement to help a human analyst, a human investigator and the like, do their job better, faster, more effectively.

MS. WILLIAMS: Thank you. My second question is for Ms. Bogen. So, how, in your view, should we be looking at the privacy implications from AI review of data as opposed to human review?

I'm thinking, for example, very popular web-based e-mail programs, famously in the past, right, scanned people's contents of e-mails to -- in order to provide them better ads. I'm told that's not done so much anymore, but it certainly will be done in the future. And I think for many people, they thought, well, if it's a computer doing it as opposed to someone reading my e-mail, then I'm okay with it.

So, how do you look at that and, you know, is it less of an issue or greater issue or the same if a computer does it as opposed to if a human does it?

MS. BOGEN: Thank you so much. There were some previous comments indicating that, you know,

existing laws and expectations should apply whether we're talking about humans or systems. But I think what's important to remember is the way in which the introduction of AI-powered systems or really any digital technology change our understanding of those processes, and whether those changes in process enable the enforcement of those laws or expectations in the same way that we intend.

So, for one, making sure that if there are rules around human access to data, is the intent behind those rules being applied in a similar manner to an AI system. But to your deeper point, I think we've moved into a world where the access to data itself is very much not the only question of privacy as we all know, it's about how that data is used and the actions that it informs.

And so, to the extent that information is being reviewed by a system that is going to inform an action that could lead to the same type of harm, either invasion of privacy in accessing information, people didn't realize was being accessed or for purposes that are disallowed, or harm that comes to

that person by the analysis of that data. It shouldn't matter whether that was by a person reviewing that data or by a system reviewing that data.

And so, the oversight can come in thinking about how is this system being actioned? What -- how are the outputs being presented to humans? And how are the outputs leading to actions that are more or less reversible? So, even if you had a human in the loop, are they empowered to do something about a system if it's behaving erroneously?

So, that sort of review of precisely what action is a system being instructed to take and how is that action triggering additional action should be the focus of oversight, regardless of where the access to data is coming throughout that process?

MS. WILLIAMS: Thank you. So, my next question is to all of our panelists. And it's actually Member Felten's question from the first panel, which I think is a very good one, which is, how do you think AI can be used to actually enhance privacy and civil liberties protections?

We always talk about the concerns about using it for other uses. But could you all share if you think, and if so, how AI can be used to improve privacy and civil liberties? And we can go in any order. I see Mr. Jaffer has his hand up, so happy to go with you first, and then then proceed to the other panelists.

MR. JAFFER: Well, I already had a chance to talk, but I'll talk very quickly about it, which is to say, I actually think there's a real opportunity here to use AI for privacy-enhancing purposes. If you think about it, programs like the metadata program, which was highly controversial and caused a lot of controversy when it was first disclosed, actually can be very privacy-enhancing in the following way.

If in fact what you're going to do to find out whether somebody is a terrorist or not, when you have a suspect number, is going to do a full content collection. If instead you're using metadata to exclude a whole set of numbers from potential content collection where you already have some amount of predication, some amount of probable cause, you can

eliminate a whole slew of people you might do very -- much more invasive collection on.

AI can play that same role by reviewing a large amount of data rapidly and vetting out a bunch of people you might do a lot deeper dive on without having to put human eyes, human hands on that data, that can be a real advantage. In the same way, you know, controversial program drone strikes and the like, right? They've actually allowed us to take much more precision strikes, a lot less casualties of civilians and the like, look, it's not perfect, but there are places where technology advances dramatically, and actually gains us benefits on the morality, the values that we have, our core values. Even though at the outset, they may seem somewhat off putting and scary, turns out when you apply them the right way, bake in, as we talked about trust, safety, and security, you can get a real advantage, actually be privacy-enhancing as long as you're not afraid of them, and slow walk the whole implementation.

MS. WILLIAMS: Thank you. Mr. Winn?

MR. WINN: Thank you. Thank you, Member

Williams. That's a wonderful question. And, yes, we should thank Ed for the -- or Member Felten for the good question.

Two things occur to me. One is obviously in connection with cybersecurity, which is an aspect of privacy that we often forget. The threats that are coming into systems from hackers are really getting to the point where the hackers are certainly deploying bots and other technologies that are in position to overwhelm humans that might otherwise try to be protecting those systems.

And so, AI has been used effectively to identify and segregate out the threat, you know, the threat attackers as opposed to the legitimate uses of access to systems. So, that's one example where more effective cybersecurity enhances a very critical privacy interest in the data being used as appropriate and not being unauthorized access.

But the other thing that occurs to me is that AI can be thought of as -- generative AI programs can be thought of as a mirror of, I mean, they're bringing out what humans are doing. And humans, as we all

know, are bias creatures, we have a lot of biases we're not often aware of, a lot of biases we're not particularly proud of, that we're often not aware of.

And so, AI can be a mirror that can be a very unflattering mirror, showing us aspects of ourselves that we don't really want to focus on. I think that -- that the sentencing controversies, the use of AI in connection with sentencing or detention decisions have brought out the unpleasant reality of the data that was being trained on was showing up the human biases that were in -- were always there.

And so, in many ways, AI can show us aspects of ourselves that are very painful, but also give us great opportunities to improve and learn from ourselves so that the ugly aspects of our characters, the failures can be then viewed as opportunities for continuous improvement.

MS. WILLIAMS: Thank you. And Ms. Garvie or Ms. Bogen?

MS. GARVIE: Sure, just to build a little bit on something that Mr. Jaffer mentioned and that's this baking in safety security. And I would add civil

liberties and civil rights into that as well, is that to the extent that we're building new tools, new AI tools or other tools right now, we have more levers to pull than if we're retroactively looking at already implemented tools.

So, from a privacy and civil liberties protection view, we actually have a unique opportunity now in the pre-implementation stage to think really critically about, can we build this stuff in by design, as opposed to can we retroactively try to build policies around already implemented tools?

MS. WILIAMS: Thank you. And, Ms. Bogen, do you think that there are ways that we can use AI proactively to protect privacy and civil liberties?

MS. BOGEN: Well, yes, I agree with Mr. Winn. I think AI, the use of AI or any technical tool can make legible decisions that were otherwise subjective or happening informally, and in that way help to -- help oversight entities, whether internal or external, identify patterns of potential misuse and correct them as well as to build in specific safeguards into the technology to protect against actions that are

otherwise disallowed.

There are other potential uses of AI, for example, to identify, you know, to a spot and redact identifiable information and datasets or to prevent the display of certain information to people who don't have access to it. But I would say those are still remain quite unreliable at this point. But there could be possibilities that AI could help play that role.

It really depends on what goal an AI system is oriented toward, and often they're oriented toward an outward goal rather than inward ones.

MS. WILLIAMS: Thank you.

MS. FRANKLIN: I think we're now over to Travis LeBlanc.

MR. LeBLANC: Thank you to everyone for joining us today and for giving us a little bit of your afternoon. I have my first question for Mr. Winn. You've mentioned appropriately, in my view, that existing laws apply to artificial intelligence systems. And, you know, generally speaking, there actually aren't exceptions in the laws that say,

except when using an AI system. How does Executive Order 12333 apply to the intelligence community's use of artificial intelligence?

MR. WINN: Well, 12333, as you know, Member LeBlanc, is a general framework for how the United States engages in its foreign intelligence work and that's generally through that executive order. There are statutes as well that are part of that framework. But it's a general framework.

The Section 2.3 of 12333 as you, I'm sure you're well aware, talks about the importance of maintaining privacy protections in connection with that activity. And the attorney general and the director of National Intelligence issue guidelines for the agencies, these are binding guidelines, for all the intelligence agencies that are developed with usually, in the last iteration of the guidelines, I'm proud to say, the PCLOB was involved in the review of those guidelines. I think that what's likely to happen in the next review of those guidelines, and I'm only speaking for myself, is that the more and more artificial intelligence systems are used, we're

probably going to see revisions of the guidelines to incorporate some of the wisdom that we've been developing to try to mitigate some of the risks and to continue the process of implementing this technology in a way that's safe and secure, that maintains the trust of the American people.

MR. LeBLANC: Okay. A follow-up question related to that, which is what kinds of revisions to the AG guidelines do you believe would be appropriate for AI applications and uses?

MR. WINN: I don't want to speculate on that question yet because we haven't started the next round of revisions. But I think the -- when -- first of all, I think the framework that was issued in 2020 by a former PCLOB attorney who has oversaw that process, Ben Huebner. And that framework really represents, I think, a really quite extraordinary forward leaning tool. And I'm -- I would hope that the national security memorandum that's about to come out, will echo many of those concerns.

But I think that those types of -- the goal is to use the technology in a way that maintains

trust, to get the benefits of the technology and to mitigate the risks. And I think that -- that the more we learn about how best to do that, and AI is sort of making us acutely aware of our ignorance. But staying aware of our ignorance is probably the best insurance policy that we have that we're not going to be deploying the technology in a way that's going to destroy trust.

MR. LeBLANC: Thank you, Mr. Winn. My next question is for Ms. Bogen.

Ms. Bogen, you identified several governance mechanisms in your opening remark that -- remarks that you believe should be put in place for responsible governance of artificial intelligence. I think much of what you've covered would largely be true of all government agencies or really any organization that is deploying artificial intelligence.

Are there any governance measures that you believe are particularly significant or should be used in the national security context?

MS. BOGEN: As you mentioned, I think the approaches to governance of AI technology are similar

to approaches of governance to organizations in general, complex systems that involve design decisions, value judgments, implementation details, et cetera. So, I would consider what have been the oversight mechanisms that have been effective in spotting potential issues in that regard as a starting point.

In general, my experience has demonstrated that there often is a significant amount of low-hanging fruit that in the excitement over the development of new technologies, tends to be its deprioritized relative to its importance, for example, basic documentation of decisions around the design of these systems, such that they can be reviewed and revisited, decisions about what data was used for the training of the system to the extent that that's shared with the government if they're procuring that system, which is a large limitation, details about what tests were run and how they were determined to be relevant to the task that was being assessed.

And details about decision -- value judgments that were made in risk management processes, while

approaches like the NIST RMF and other mechanisms are fantastic at helping to structure the design and development and review process. They still leave on the table many open questions around how you weigh the information that is revealed throughout that risk management process against the goals of an organization. And those are where value judgments come into play and where sometimes, unfortunately, civil rights and civil liberties end up falling below the line or at least lower than we would like.

So, anything that can enable the spotting and review of those types of decisions will support beneficial scrutiny of their development and deployment in the long-term. And they can facilitate the building of other governance mechanisms on top of that. But without that foundation, it will be very difficult to build other effective governance mechanisms.

MR. LeBLANC: Thank you. And Ms. Bogen, do you believe that there are any AI applications that should not be used in the national security context?

MS. BOGEN: My response, there will be any AI

system that is performing a task that we would not want an intelligence mechanism to be doing at all, for example, real time facial recognition, whether that was a human who was very good at recognizing people or an AI system, we wouldn't want pseudoscientific goals, for example, emotion recognition or other things that come to mind.

So, being mindful of what are the parameters around which a system is oriented and do those fit in to the overall structure and values of the organization that is deploying the system.

MR. LeBLANC: Okay. Thank you. Next, I want to ask a question to Mr. Jaffer. You discussed in substantial detail in your opening remarks about terrorism and in particular the foreign connection to terrorism. But, you know, as I'm sure you're aware, terrorism is also a domestic threat. It's not just a foreign threat. And in fact, domestic terrorism is the number one terrorist threat to the United States, not foreign terrorism, although many may be surprised to learn that.

The key privacy and civil liberties challenge

in the domestic context is that the government is usually looking at U.S. persons or somehow obtaining through collection, the information from or about U.S. persons. And I fully agree with you that there is a need for the government, if it's going to deploy AI, to build in trust, safety, and security.

The fundamental problem that the government and, in particular, the FBI, since you're referring to the FBI director, the fundamental problem they've had in the past is a lack of trust. And so, how can the FBI build trust that its access to massive troves of data about U.S. persons will not be processed through AI systems in ways that are inconsistent with current norms?

And are there any limitations or safeguards that you believe should be put in place to protect against AI abuses by the FBI? For example, should the FBI be able to use artificial intelligence to predict who is or may be a criminal?

MR. JAFFER: These are great questions, Member LeBlanc. I would say, let me start with at the end of your last question first, which is to say, no,

I don't think we want sort of a predictive system predicting who are criminals. It sort of reminds me of that, the movie whose name I'm going to forget, Minority Report. And I don't think anybody's looking to sort of embody a minority report system at the bureau, whether it was highly trusted, which it used to be back in the past or is less trusted today.

And as you know, the trust of the FBI has waxed and waned over time back in the post '60s and '70s era when there were the days of the counterintelligence program, Operation CHAOS, the CIA. There was a deep mistrust of the FBI. And we put in place a lot of policies and procedures to address those and bring them back into a more positive light.

I think we've seen a decay in that trust in the more recent era as well, in part because of situations that we've seen in both political parties, as well as the popular dimension where there's been a decay in trust in not just the FBI, but all of our law enforcement and rule of law institutions, including the Justice Department. And that's been a real challenge.

It's been in part, I think, fomented by overseas actors as well, but there is some -- there are legitimate reasons for some of that distrust. And you referred to some of them, some of the challenges we've seen in programs like 702 and 215, where as a general matter, the FBI has been doing a very good job, but they make errors. They make mistakes. The mistakes sometimes are of large scale.

And so, then they self-report these mistakes to the FISA court, and then it turns into this large issue of, look, the FBI is violating privacy and civil liberties, when in fact they're identifying errors they made. Yes, there are mistakes. They're not intentional. The number of intentional violations are very, very few, whether at the FBI or the NSA.

And so, what we don't have is an epidemic or a pandemic or any sort of demic of intentional violations of private and civil liberties, but a lot of mistakes and a lot of errors and that erodes trust. You're right to say that. And so, the question then is, how do you rebuild that trust? And that's going to be a challenge. It's going to be a challenge as we

deploy tools that are more and more capable, more and more capable of taking in large amounts of data and processing it quickly.

I think what we have to understand is the more data you take in, the more data you process, the more likely you are to make mistakes. The question is, what do you do about those mistakes when you make them? Do you put in place policies and procedures as we've done with the bureau, as we've done with other agencies, like the attorney general guidelines, to guide those and to fix those?

And how often do you self-report those? How often do you get caught making an error or get caught making an intentional problem? And where there's an intent and there's an actual failure where somebody's done something wrong, do you throw the book at them?

We had an example, you know, a famous example of a lawyer who lied to a court to obtain a FISA, right? Changed a material fact. That guy got time served, right? He didn't get time served. He got probation. That is crazy. That guy should have gone to jail for a long time. He should have been stripped

of his license. I understand he did lose his license for a while. He got it back. That is unacceptable.

When people make failures in the FISA context where you've got ex parte and in-camera proceedings, you have to throw the book at them. Otherwise, that and all the unintentional mistakes get all bottled together. And we have a situation where fundamental trust is undermined.

And I think that's an important role that Privacy and Civil Liberties Oversight Board can play when putting out reports like this one on 702 where they're really -- where you're really candid, right, and very clear about intentional versus unintentional mistakes and not sort of combine the two and treat them like they're the same thing because they're not the same thing.

MR. LeBLANC: Thank you very much. My time's up. So, I'm going to go ahead and pass the microphone on to Member Felten.

MR. FELTEN: Thank you. I'd like to ask a question to all of the panelists related to algorithmic bias. And I'll ask it in the context of

facial recognition, which is an area where perhaps we have the best and most extensive data from NIST studies. And in this respect, we see two things happening at the same time. First, we see that according to NIST studies, the demographic differentials in error rate of the very best algorithms are shrinking considerably over time.

But on the other hand, we see continued instances of harm to individuals due to, for example, false arrests in a pattern that is very obviously correlated with race. And so, there's some gap between what the algorithms can do and the results that we're getting in the field in this area.

So, I'd just like to ask the panelists, you know, if you could talk about what may be happening there and in particular what we might do to address this disparity so that at least we can reduce the level of errors closer to what the algorithms can provide.

And let me go through the panelists in alphabetical order starting with Miranda Bogen.

MS. BOGEN: Thank you, member Felten. I

think when we think about algorithmic bias, there are a number of different lenses through which to consider that. So, one is simply comparative performance of a model itself vis-a-vis its specific goal. Facial recognition, for example, has a very particular mechanism by which to measure if it recognized a specific individual. And by disaggregating that measurement across demographic groups, you can identify if there are those disparities.

Running those tests is one way to identify where the gaps are and facilitate attention to closing them. A way to continue making progress in that front is considering what are the groups by which the measurements are disaggregated. Are those salient to the errors that are being made or are there additional disaggregations that would illuminate the causes of those gaps, which may or may not, and likely are, but may not be fully correlated with legally protected groups.

So, another approach to conducting these measurements, in addition to disaggregating by predefined groups, is identifying clusters of errors

of systems and trying to consider what might be driving those errors by reviewing those errors. That's to address technical bias in that way, disparate performance against a metric.

But in, you know, similar systems, there are also questions around, was the goal of the system appropriately defined? Was the target metric reflective of some kind of underlying assumption in the world that incorporates some kind of historical bias against which disaggregating measurements of the system would not reveal and requires considering a system more holistically?

And in other cases, systems might reveal biases that are simply reflected in the world where technical intervention are not the most opportune approach to address that, but rather reflecting on overall processes. So, I would divide it in that way.

MR. FELTEN: Thanks. Clare Garvie?

MS. GARVIE: Thank you for the question. I think I have two points on this. One is the operational conditions point, and that is that NIST, while the tests that NIST performs on face recognition

are extremely valuable, they still don't represent what happens in operational conditions. And that is a sociotechnical system, a series of steps for which the algorithm is one of multiple steps.

So, until we have actual testing on face recognition in operational conditions, whether that's in the law enforcement or a national security standpoint, I think we are going to have these differences in what the tests show in terms of accuracy, reliability, and bias, and what we're seeing on the ground in practice. For example, the human-in-the-loop, is that a valuable check against misidentification or does the cross-race bias effect actually exacerbate or perpetuate the racial bias in a way that isn't being tested by NIST?

And the other point I would raise is that, yes, while the differential error rates across race have declined over the last few years in the top performing algorithms, it does seem that race, sex, and age still impact the accuracy measurements or the reliability scores given to non-mated pairs.

And my suspicion here, sorry to get a little

wonky on this, is that the algorithms are still confusing class characteristics in individual characteristics, that, yes, algorithms are not looking for race, sex, and age, but they are looking at race, sex, and age to determine individual identity, but those are class characteristics and not individual ones.

So, the types of mistakes that an algorithm is going to make is going to be between people who look very similar, aka people of the same race, sex, and age, which again leads us to question whether the human in the loop is actually performing a valuable check, when the algorithm is making the same types of mistakes that humans are going to make, which is confusing people who are in the same demographic cohort.

And then we put that all into a system of law enforcement in the examples of face recognition mistakes that we have, and we have a system that overwhelmingly over-investigates and over-incarcerates particularly young black men. So, the system is going to not operate independently of those existing biases.

MR. FELTEN: Thanks, Jamil Jaffer?

MR. JAFFER: Yeah, look, I think both Ms. Bogen and Ms. Garvie have laid out a great set of examples and things that we might do to address some of the challenges that we face in this domain. You know, I think at the end of the day, one thing that ought to be considered is that, you know, these systems are designed to mimic human reasoning, right? They're designed to function as neural networks that connect various aspects of disparate information to create a holistic picture the way the human brain does. So, it's not surprising that they're going to have some of the similar hiccups that human brains make, whether based on intentional or unconscious or other forms of bias or other cognitive errors that a human brain makes.

In some ways, the design of a neural network is designed to do that. And I do worry that we have in the AI domain the same kind of fetish that we have in the cybersecurity domain, which is that because we're talking about zeros and ones, we expect perfection, right? That's not realistic in the

cybersecurity domain. It's not realistic in the AI domain that we will not get to perfect. We will get better and we're seeing that in the results over time, but if we expect perfection, we'll never be satisfied.

And so, you're going to have to have some amount of human involvement, human judgment layered on top of, and admittedly, to Ms. Garvie's point, admittedly deeply flawed human judgment and sometimes biased human judgment layered in on top of a potentially biased algorithm that's designed to mimic a human brain.

And then the last piece of it is, the data we're feeding in to train these models has its own biases built in depending on how you build the data and how you address the data set. So, you can bake some of that out as well. You're, again, not going to get to perfection, even though these are zeros and ones, and these are computers, you cannot expect perfection. You'll be disappointed every time, and there is going to be some level (inaudible). The best thing you could do is to try to train out and then ultimately layer in human judgment and recognize both

human judgment and computer judgment are never going to get you the result that you ultimately want. You can only get better, not perfect.

MR. FELTEN: Thanks. Mr. Winn?

MR. WINN: Thank you for the question, Member Felten. I'm as troubled as you are about why the -- you just continue to see abusive patterns involving facial recognition, even though the algorithms are getting better. I think it's because of a simple mistake. And I think to some extent, Jamil pointed it out and Clare and Miranda also suggested it as well.

People are using AI facial recognition tools because they've been watching too much TV and they see the word match come back. AI facial recognition tools have never been intended to be used to create a match. You know, for instance, if you have a chance to look at the privacy impact assessment that was done for the FBI's use of facial recognition technology in connection with the NCIC, FBI CJIS programs, what you'll see is the FBI never permits a single photo to be given to an investigator. You get an array, and that array doesn't go to the investigator until an

independent group of trained -- people who've been trained in biometrics review the array.

The investigator is trained never to rely solely on the photo or their own judgment about who in the array is most likely to be the suspect, but to seek corroborated evidence. And those, you know, again, Dean pointed out earlier today that it's not simply the machine and it's not simply the human, but the rules relating to the interface between the machine and the human.

And you've got bad rules at the state and local, and it's mostly a state and local problem, bad rules about how to implement and use that technology against a background of a lot of biases that we humans tend to have, against a framework where you're thinking of this tool as being a silver bullet. Law enforcement has always been about putting together corroborating evidence to reduce the level of uncertainty, not to achieve some perfection.

And I think the reason you're continuing to see these instances is because people are misunderstanding what the tool can do and should be

used for -- can be used. And they're not following best practices that have been developed. And I want to give the FBI a shout out on this case. They've really developed some excellent tools that many privacy and civil liberties organizations have championed because they have shown it's the human machine interface via these operational rules that are constantly being evolving and improving that ensure that you can use this technology in a very reliable, trustworthy way where you don't destroy trust.

MR. FELTEN: Thanks very much. Let me pass to Chair Franklin.

MS. FRANKLIN: Thank you. So, I want to start by building a little bit on that last round of questions with member Ed Felten and turn to Clare Garvie. So, you have done a lot of research, I know, on facial recognition, particularly in the context of law enforcement. And in your opening statement, you addressed some of the risks presented by the use of AI for predictive purposes. And in particular, you noted that many AI tools over-promise on their ability to predict a propensity to commit violence or to identify

threatening behavior.

So, my question to you is, whether there are any best practices that you would recommend for any government agency, you know, broadening out from the facial recognition context potentially, but any best practices for a government agency seeking to use AI to conduct any type of pattern analysis for predictive purposes in the counterterrorism space? And what kinds of safeguards would you want to incorporate?

MS. GARVIE: Thank you so much for that question. I was struck by one of the questions to the previous panels about what is the most serious or important aspect of this vast space of AI to focus on and how does the board choose that? And it did get me thinking that the predictive space of algorithms does seem to be one of the most critical in terms of focus because of this changing who we view to be the end arbiter of a decision of maybe it is anomalous behavior, but maybe it's suspicious behavior, maybe it is behavior that rises to the level of probable cause to form an interdiction or to take negative action against somebody. I do think this is where AI maybe

runs the risk of having the greatest harms, this sort of outsourcing of guilt, if you will.

I think there are a couple of different mechanisms to approach these tools. One is I think a go, no-go analysis. Is this a place where we want to automate decision-making or do the harms of that decision-making or the mistakes that that algorithmic system might make outweigh the benefits of moving from a human or more cautious, slow approach to an AI-based approach? So, that's one analysis to do before the implementation of a system.

I think the next one is, okay, is it reliable? Does it do what it says it does? And I think there's far too little engagement with this question before we implement advanced automated tools, particularly AI in the law enforcement and other spaces. We do have this inclination to see that AI can solve mass data problems and then we implement it without analyzing. Is it reliable?

So, I think that's the next check. Does it reliably do what it says it does? And can we get it there? Or are there -- does the human in the loop

solve the problem or not? I think of humans in the loop as being necessary, but maybe not sufficient to answer a lot of these AI questions. Just sticking a human in the loop may actually exacerbate reliability problems and we have to be very cautious around that as a mechanism, but it is certainly a mechanism to -- or a lever to pull, if you will. One.

Another is privacy by design. What data are these systems operating on? Does it have U.S. person's data? I think this has to be really carefully evaluated with DHS use on soft targets versus maybe intelligence use abroad. DHS has now partnered with Analytical AI to do anomalous event detection on soft targets. I think there are very real questions about whether it's appropriate for an AI system to be determining what constitutes anomalous or suspicious behavior at a stadium, for example. So, what the target is, I think, is another mechanism, another area where you have a moment to decide this cost-benefit analysis.

And then I would look to the -- this existing privacy impact assessments and systems of records

notices. And really urging, and this is not unique, I'm not coming up with this idea, the National Security Commission on the AI recommended this. AI moves extremely quickly and it manifests, they changes the way systems and databases operate. PIAs maybe need to keep pace with that. It's not really sufficient for the automated targeting system to have a PIA from 2017 if it's using AI systems from, let's say, 2022 and beyond. So, I think there are a number of mechanisms throughout the lifecycle of developing and then deploying systems that I think we need to think very carefully about and pull each and every one of them, depending on the harms identified.

MS. FRANKLIN: Thank you. Okay. I want to turn next -- oh, I see fingers, I hate seeing hands. Now I'm going to turn next to Miranda Bogen. So, if I still have time, I'm going to give that question, and then I'm going to treat those as two fingers coming back to Clare Garvie. But since I'm going last in this round, I'm going to turn next to Miranda Bogen with a question and I'll come back to others who raised their hands if I have time.

So, you've written about the many risks of AI use and recommended that employers and others take active steps to detect and mitigate or remove bias in their systems. And in your opening remarks, you spoke about the problems that can result from incomplete, unrepresentative, or biased training data. What risk mitigation tools do you recommend to avoid those harms or to detect them before an AI system is implemented? So, there's been talk about things like risk assessments, whether those in your view have been effective in testing algorithms before they're deployed, and audits, which, I guess, can be either before or after the fact? And what do you think is the best fit for government's use of AI in the counterterrorism space?

MS. BOGEN: Under any -- whatever label people would like to call them, whether risk assessments, audits, impact assessments, any proactive step to review and test systems before and after deployment will help to identify more issues than not conducting those tests. And unfortunately, too often, those tests are not conducted either before or in an

ongoing manner.

It's very important to test systems before they're deployed because there could be a number of design choices or even different versions of algorithms that have the same quantitative results that an institution might be evaluating success along, but significantly different patterns of errors within those results. And so, by disaggregating those measurements along groups of interest, protected characteristics, or other vulnerable groups, there can be a comparison done to say, in pursuit of a goal that we may find to be reasonable, which version of a model or a system that incorporates that model best accomplishes that goal while resulting in the least likely harm.

And I think previous folks have said, you know, you can't entirely remove all bias from systems, there are many sources of that bias; data missingness, you know, assumptions about what data is relevant, that may be more pertinent for one population than for another, et cetera. But by doing that type of proactive measurement, you can spot whether those

assumptions might have led to disparities that would be of concern.

And then continuing to conduct those measures on an ongoing basis is important because, as Ms. Garvie said, the conditions of deployment may significantly differ than the conditions of testing. And so, unless you're testing that in the wild and also understanding how human are acting on the output of systems to the extent that they're being relied upon to do so. We won't know if -- even if there might be no disparities in the performance of the system, which is highly unlikely, but even if there were, whether the humans are acting differently in similar circumstances in a way that would need to be identified and for which processes would need to be implemented to prevent that from happening.

So, again, very simple approaches, but alas, don't tend to be prioritized across the board. And so, whether they're incorporated into impact assessment, risk assessment, audits, risk management processes, those types of considerations are necessary alongside considerations of overall accuracy of a

system independently.

MS. FRANKLIN: Thanks. Okay. I think I have 2 minutes left before our final lightning round among back to the board members. So, I just want to give a chance to Peter Winn and Jamil Jaffer, you know, super quick, what you raised your hands for just before?

MR. WINN: Just the point of the -- doing privacy impact assessments after the system and the operational procedures have been established is useless. You have to do them early -- during the early development phase of the process. And then, what inevitably happens is, you know, you're trying to predict as much stuff as you can, you can't predict everything. I look back on privacy impact assessments I signed in 2017 and I'm appalled today at all the things I've missed. So, you have -- it's a continuous process, you have to have those privacy impact assessments going back and looking at what you now know and, you know, and then you do it again because otherwise, you're really going to lose the benefit of all the knowledge that you're gaining through, you know, your ignorance. You know, you're mitigating

your ignorance and you're not -- you have to go back and cycle this risk mitigation process. It's a continuous --

MS. FRANKLIN: Thank you. I'm just giving 30 seconds to Jamil Jaffer for what he was raising his hand for.

MR. JAFFER: You know, look, I think Ms. Bogen, Ms. Garvie, and Mr. Winn said almost exactly the same thing, which is that you got to do this from the beginning, bake it in from the beginning and then to do all the way through. But the key is, if you're just the government doing and saying, we're going to do reassessments, it's not going to work, you've got to incentivize industry players and investors who are already incentivized in their own ways to do these things. And the core there is this idea that trust, safety, and security actually benefits your return on investment, it benefits the intellectual property that you're creating, it benefits the uptake of these capabilities. And the more the government can meet that part of their buying mechanism, and part of their feedback to industry and investors, that's really

what's going to drive this thing. It's not going to happen because you do a bunch of PIAs over in the government. The key is to bake this in and have a continuous assessment process go on.

And heavy-handed regulation by the way, is simply going to suppress innovation. What you really want is incentivization of the right kind, but also gives you the answer that we all want, which I think everybody agrees on.

MS. FRANKLIN: Okay. Thanks. Okay. So, final lightning round back up to Board Member Beth Williams.

MS. WILLIAMS: Great. Thank you. So, one of our former board members, Jim Dempsey, has written extensively on the importance of contestability in AI systems. And so, I'm wondering if in this lightning round, you can just quickly tell us if you think contestability is an important consideration and how you think we can best incorporate contestability into some of these systems. And so, I think because I'm a Williams, I'm going to ask to go in reverse alphabetical order, starting with Peter Winn.

MR. WINN: Well, thanks. Jim Dempsey is one of my favorite PCLOB board members and currently one of my favorite privacy, data protection review court judges. So, he always has very thoughtful things to say. I think that contestability ideas that he has also involve asking the right questions and probing in the right ways, and a multi-stakeholder process. And that's also described in the intelligence community's framework for IC development.

The importance of having multi-stakeholders engaged in pushing and asking those questions from lots of different perspectives because none of us have that monopoly or knowledge that we all wish we had. And bringing in that multi-stakeholder process to the extent you can and you can -- even in a classified environment, you can bring in a lot of multi-stakeholders. The PCLOB itself represents a very diverse body of board members representing, you know, a similar kind of diversity of views and judgment, pressing all of the aspects of the development of these programs. Looking at the underlying data, understanding how the algorithms work, all of that is

critical. But it can't be done by a single point of view. It has to be done in a multi-stakeholder way.

MS. WILLIAMS: Thank you. Mr. Jaffer?

MR. JAFFER: Yeah, look, I think, obviously contestability and being able to push back against a decision made by AI for an individual is the right thing to do. I can't imagine anybody on this panel is going to disagree that you should have contestability baked in. And so, to me, you know, the Dempsey, Landau idea of contestability is exactly the right one. I do want to say something to the data protection court, which is that (inaudible) contestability in America -- in the American system, it should be for Americans. This idea that we're bringing Europeans in, and we're giving them this fake court made up of executive branch officials is ridiculous, and completely antithetical to our system and makes no sense whatsoever. So, I did want to put that out there. I do love Jim Dempsey, Data Protection Court, ridiculous.

MS. WILLIAMS: Thank you. Ms. Garvie?

MS. GARVIE: Yes, just echoing

contestability, super important on -- in at least two aspects. One is the foundational validity. Does this work as intended? Have we tested it? And have we tested it sufficiently and independently? And then the validity as applied aspect as well. If it didn't go right or wrong in this particular case, and can the person directly affected, challenge that and contest that?

MS. WILLIAMS: Thanks. Ms. Bogen?

MS. BOGEN: Yes, I was honored to participate in the series of workshops that led to the report on contestability by Mr. Dempsey and Ms. Landau. And so, I would certainly endorse the findings of that workshop. And I think one of the main conclusions that came out of that is contestability is not an independent concept from due process. And so, we need to remember all of the circumstances under which due processes is guaranteed, and the introduction of AI does not change that. I think we also need to be attentive to the limitations of these systems and the justifications they can or can't make around the recommendations and ensure that human analysts, again,

similar to the case of facial recognition, don't solely rely on the output of AI systems to justify actions that would otherwise not be justified without corroborating evidence or signals.

MS. WILLIAMS: Now, turn it over to Member LeBlanc.

MR. LeBLANC: Thank you, Beth. Mr. Winn, I hesitate to suggest that I might also have read some of your early PIAs and wondered a few times what you were thinking at that time. And I won't ask you to tell us which ones of those are outdated so that you can correct them. But I do want to go back to the question that I had posed to Mr. Jaffer in the last round, because I did notice that you came off mute right after he finished his comments. And it was about the FBI's prior errors in the FISA context. And of course, recognizing that the bureau is a component of the Department of Justice, it's only fair to give you an opportunity to share any remarks about how the FBI can build trust in its use of AI.

MR. WINN: Thanks -- thank you, Member. I'm very grateful for the opportunity to respond. And

we'll have a separate conversation about which ones -- which PIAs that I signed in the past are most in need of updating. The -- but, you know, we've been talking about the machine, we've been talking about the human, and we've been talking probably not enough about the interface. And the interface being the rules that apply when you're deploying the human and the machine to accomplish a mission. And your question earlier about 12333, the guidelines, the attorney general guidelines that are issued pursuant to 12333 and at the FBI, I would highlight how they in turn implement the attorney general guidelines through the DIOG or the Domestic Investigations Operation Guide. That's a massive standards for good law enforcement, okay?

I would say that, you know, Jamil was talking about the difference between accidents and on purpose. And even a dog knows the difference between being tripped over and kicked. When the FBI -- and the Durham report which was issued by Special Counsel Durham, discussing a breakdown in trust at the FBI. If you read it carefully, you can see that what he points out is the intentional violation of their own

rules, the DIOG. Now, the DIOG isn't required by statute, isn't a regulation, but it is an -- for years, it was the FBI's Bible. It was what made you an FBI agent and why the FBI was so much better than any other law enforcement agency at what it's -- it was a self -- it was the identity of what it meant to be an FBI agent. And the DIOG was established by -- originally by Attorney General Edward Levi to deal with the breakdown of trust that took place under Edward -- you know, Director Hoover with the COINTELPRO scandals where they were going up on Dr. Martin Luther King, you know, based on not enough evidence.

And the DIOG was put in place with thresholds of evidence, you can't open an investigation with just a little evidence. You can do an assessment, you have to do a preliminary investigation and only when you get more evidence, then can you open up full investigation. And only when you have a full investigation, can you actually get a wiretap or something like that. That's -- those things are baked in. They are the interface. They are the rules that

don't go away when we bring in AI. And the mistake that people are making is they're thinking those rules don't apply anymore, then we do our --

MS. FRANKLIN: Thank you. Peter? I'm so --

MR. WINN: -- we're checking to make sure those rules are still being followed, and trust is maintained. And so, that's come to a conclusion.

MS. FRANKLIN: Thank you. Sorry, just trying to make sure that Ed Felten and I get a chance for our last lightning round question. So, over to Ed Felten.

MR. FELTEN: Okay. Yes, I want to come back to a question that Mr. LeBlanc asked Ms. Bogen earlier. And that is about how the general frameworks for AI governance, such as the NIST framework might need to be adjusted or augmented in the context of counterterrorism. And in the interest of lightning round efficiency, I will ask each of the other three panelists who have not yet addressed the question to give a brief answer on that topic. And I'll start with Mr. Winn.

MR. WINN: So, the brief answer is we've got to focus not simply on doing the general risk

assessments that NIST is talking about. But we have to make them context specific. And we have to look at our prior rules of engagement, the rules that have been developed through knowledge and experience of generations of law enforcement and national security individuals. Those rules need to be baked in as well to the risk assessment process.

MR. FELTEN: Thanks. Mr. Jaffer?

MS. JAFFER: Yeah, I mean, look, the AI risk assessment NIST, you know, frameworks are frameworks. They're not designed to be the exact thing you implement every day, day to day in and out. They're designed to be customizable to a variety of contexts. And so, I think in the government context, you ought to apply them in a way that makes sense. And that accounts for the unique issues that Peter and Ms. Bogen and Ms. -- and the other panelists have -- and Ms. Garvie have raised as well. So, I think just -- you got to apply the frameworks in a contextual way. So, I don't think there's anything surprising there.

MR. FELTEN: Right. Ms. Garvie?

MS. GARVIE: Agree. And I would also maybe,

this is pie in the sky, but I would love to see the intelligence and national security community also adopt something that DARPA is now adopting from the genomics project, which is broadening it a little bit, the risk framework a little bit out to ethical, social, and legal implications. So, a little bit broader than just privacy because I think that helps anticipate potential future problems or challenges and concerns caused by AI systems to the point that we've been talking about earlier with these PIAs getting out of date so quickly. And just as an example of how we need to anticipate from a broader perspective, the implications of these systems.

MR. FELTEN: Great. Thank you to all the panelists. And I'll pass to Chair Franklin.

MS. FRANKLIN: Thank you. Okay. So, final question to close us out. If you can, each, I'm going to go through in forward alphabetical order again. If you can each give us a concise framing as you can of how would you scope and define an appropriate slice or focus for PCLOB's oversight of AI in counterterrorism? So, starting with Miranda Bogen.

MS. BOGEN: I'm sure the other panelists will have very insightful perspectives on the question itself. So, I will just say no matter what slice PCLOB chooses to focus on, you should also make recommendations about to the extent there are other elements that PCLOB is not going to focus on or it isn't within their ambit to, that other analogous oversight mechanisms are set up to focus on those other elements.

MS. FRANKLIN: Thank you. Clare Garvie?

MR. GARVIE: I think it's less a question of what the appropriate slices and more sort of a hierarchy. But my current hierarchy in the course of this conversation, I think would place predictive systems where we've supplanted human decision making with an automated decision sort of at the top of that hierarchy, followed probably by systems where AI is being implemented into an existing structures, particularly mass datasets, and fundamentally changing the nature of the data and its applications where the impact assessment has been already assessed at collection, but AI is changing the applicability,

usability of that data into something that's new and raises new challenges.

MS. FRANKLIN: Thank you. Jamil Jaffer?

MR. JAFFER: Yeah, I think the key focuses for PCLOB has to be staying within a statutory mandate, right, which is about efforts to protect the nation against the threat of terrorism, right? What you don't want to do is end up with the PCLOB on a roving search for AI challenges with national security more generally, right? What if the PCLOB's mandate was broader, right, that's a different question for Congress to consider if they want to broaden your statute. But to the extent that they've given you a statue you have, you've got to stay within the counterterrorism construct. And so, to the extent that AI is being used in the counterterrorism mission space, or is going to be used in that space, that's a place for PCLOB to focus. I don't think there's necessarily a specific slice within that. But staying focused on the counterterrorism mission, and not getting into the related national security matters, I think will be the thing that allows PCLOB to do its

job most effectively.

MS. FRANKLIN: Before we go to Peter Winn, I just have to say, our jurisdiction clearly covers multiple purpose programs and activities that include counterterrorism with Section 702 being --

MR. JAFFER: We can debate that. We can debate that. I'm not sure that's the right way of the statute.

MS. FRANKLIN: Over to Peter Winn.

MR. WINN: I'm not going to get into the jurisdictional debate. But I know that as Member LeBlanc said, domestic terrorism is a serious concern. And the domestic terrorism context in the United States is done through the law enforcement structures. The law enforcement structures have well developed rules, like the FBI DIOG. I would look at this question, which is, is the law on the books, law on the ground? Are people actually doing what they say they do? And if they're not, you need to hold us accountable. And the PCLOB is in a special position to make transparent, both the things that we're doing right as well as the things that we're doing wrong, so

the public better understands how to evaluate whether this is done on purpose or whether it was an accident.

MS. FRANKLIN: Okay. Thank you all and thanks to all our panelists for sharing your insights with us today. And thank you to everybody who has been joining in our audience and this will close us out. Thank you.