Center for Strategic and International Studies

TRANSCRIPT

Event

"We Hold These Truths: How Verified Content Defends Democracies"

DATE

Monday, March 27, 2023 at 9:00 a.m. ET

FEATURING

Andy Parsons
Senior Director, Content Authenticity Initiative, Adobe

Mounir Ibrahim
Vice President of Public Affairs and Impact, Truepic

Andrew Jenks
Principal Program Manager, Azure Media Security, Microsoft, and Chairperson, Coalition for Content Provenance and Authenticity

Beth Van Schaack
Ambassador-at-Large for Global Criminal Justice, Office of Global Criminal Justice, Department of State

Jessica Brandt
Policy Director, AI and Emerging Technology Initiative, Brookings Institution

Matthew Turek
Deputy Director, Information Innovation Office, DARPA

Dana Rao
Chief Trust Officer and General Counsel, Adobe

CSIS EXPERTS

Emily Harding
Senior Fellow and Deputy Director, International Security Program, CSIS

Suzanne Spaulding
Senior Advisor for Homeland Security, International Security Program, CSIS

Transcript By
Superior Transcriptions LLC
www.superiortranscriptions.com

Emily Harding: Good morning. Thank you so much for joining us today for a critically important discussion about deep fakes, democracy, and simple ways to verify the truth.

This week the Summit for Democracy will hold a series of discussions about defending democratic institution against an onslaught of modern threats. Authoritarian regimes know they cannot compete with the core idea of personal freedoms. Instead, they seek to sow doubt: doubt in each other, doubt in our leaders, doubt in our institutions, everything about the way our society functions. We will not let them succeed, but stopping the slide towards the end of truth means building trust, trust in what we see and hear and in the process of democracy itself.

We are pleased to host this important event, alongside the summit, in support of the larger goal: defending our democracy against those who would seek to make us doubt its strength. It's a twofer at CSIS: Tomorrow my colleague Marti Flacks will host an event on business and democracy, including concrete strategies that forward-thinking private sector leaders are advancing to strengthen democracy around the world.

We entitled this event "We Hold These Truths" not only because those are the opening lines of a document central to our democracy, but because the very concept of truth is under attack. Today we will discuss the challenge that deep fakes, cheap fakes, and generative AI pose to a society that depends on an engaged, invested population to create a healthy democracy. Thanks to this technology, we can do silly things. We can put Ryan Reynolds' face on anybody. Movies like "Free Guy" can make movie magic using real AI inside a movie about AI. We can also do terrifying things. An early deep fake of President Obama showed us that putting words in the mouth of the leader of the free world was not only possible, it was done. In the early days of the Ukraine war, a similar stunt was attempted with President Zelensky. The video showed the Ukrainian president telling his troops to stand down, to surrender, to lay down their arms. Luckily, it was a terrible deep fake. But we will not always be so lucky. One of our researchers has conducted a study showing that your odds of identifying a deep fake are roughly 50/50. You'd be just as good flipping a coin.

So how to address this problem? A portion of today's event will discuss one option. A coalition of companies that produce tech solutions has come up with a way to show the entire chain of custody of an image or a video, creating confidence that the image is a true representation of what happened. Now, today's event is not meant to be an advertisement for any one way to verify content. This is one option, and I hope it will be one option among many for reassuring consumers that Ryan Reynolds is indeed the real Ryan Reynolds, or more importantly, I am the real Emily Harding,

and later you will see the real Suzanne Spaulding hosting a real panel with real experts on where democracy and technology meet.

Today we have a member of our staff walking around taking photos with one of the technologies we will discuss here today. Those images will appear on our website shortly, including the icon that will let you understand the provenance of that image, from my colleagues' hands to your screen.

Today's order of march is as follows. First, we will invite some of those forward-leaning private sector partners up on stage to discuss their approach to ensuring content provenance. Then my colleague Suzanne Spaulding, who is basically the Captain Marvel of defending democracies, will moderate an all-star panel, including Ambassador Beth Van Schaack, the ambassador-at-large for Global Criminal Justice at State; Jessica Brandt, the policy director for the AI and Emerging Technology Initiative at Brookings; and Dr. Matt Turek, the deputy director of the Information Innovation Office at DARPA. Finally, Dana Rao, the general counsel at Adobe, will speak about how businesses can help defend democracy with initiatives such as C2PA.

It's a packed agenda so I will hand it over to our first guests. They are: Andrew Jenks, who is the chairperson of C2PA. As chairperson he leads a cross-company team of experts in developing the next evolution of authentic media. He is also a principal program manager for Microsoft Azure Media Security, and the director for Microsoft's media provenance investments across the company. Andrew has spent more than 30 years in the computer science industry creating, incubating, and releasing new solutions to complex problems.

Andy Parsons is the senior director of Adobe's Content Authenticity Initiative, which is creating open technologies for a future of verifiably authentic content. Prior to joining Adobe, Andy founded Workframe, served as chief technology officer at McKinsey Academy, and co-founded Happify, a mobile platform for digital therapeutics and behavioral health.

Mounir Ibrahim is vice president of public affairs and impact for Truepic, which specializes in digital content provenance and authenticity. From 2009 to 2017, Mounir was a Foreign Service officer with the U.S. Department of State and a key Syria adviser to various ambassadors and presidential cabinet members.

Mounir served in Damascus, Washington, Istanbul, Bogota, and New York at the U.S. Mission to the United Nations. He's also an adjunct professor at Columbia and a TEDx speaker. So we look forward to be entertained.

Over to Mounir, Andy, and Andrew. Thank you.

Andrew Jenks: Thank you, Emily. I'm Andrew Jenks and I'm – it's a pleasure to be here today joining CSIS for this august series of panels and speakers.

Why is the Coalition for Content Provenance and Authenticity here today? Since our founding three years ago we've been focused on issues surrounding media and transparency. Of course, today's discussion is one about trust. But transparency underlies that trust. If you don't know where the media that you're consuming came from you don't know whether you consider it trustworthy.

We get trust and transparency confused all the time because they're very similar concepts. But I like to think of the fact that trust flows from the user backwards. Newspapers don't choose who trust them. I choose who I trust.

What I deserve, though – what I deserve from all of the news sources that I receive is the transparency in the media so that I know that it actually came from the news source that I have made the personal choice to trust.

That's what the C2PA is focused on, providing that transparency. When we released our first specification in December of 2021 we were focused on images and video and providing provenance and transparency specifications for those.

Since that time we've grown. But what is this notion of provenance? What is media provenance? It is, at its core, this idea that you can bind basic facts to a piece of media in a way that's unforgeable so that you have the ability to say, for example, this piece of media was published by a news network, this piece of media was captured at this particular location, or any other information that you want to provide along with it. That information can travel along with that media so that when a consumer receives it they can see where it's been, where it came from, what its origin was.

Since we released our first specification we've been continuing to grow and progress the technology that C2PA supports, moving into audio files, moving into different file types, and most recently in our 1.3 specification, which will be released next month, moving into describing how generative AI content is created.

We think this is hugely important because we can't simply label material that is authentic. It's important to be able to discern information from well-meaning actors that is intended to be inauthentic. How do we deal with parody? How do we deal with satire? How do we deal with creative expression?

C2PA and its 1.3 specification gives us the capability to explain these and other concepts. The C2PA is not just a technical play. We have membership from NGOs, from CDN providers, from news organizations, from press organizations and others all across the media spectrum.

We believe this is necessary because mitigating deep fakes isn't going to take simply the technique that we've produced or any other technique. It's going to take an ecosystem wide solution for an ecosystem wide problem.

When we first introduced this technology we stood on stage and said, this is how Mecklenburg County, a local county government, is using provenance technology today. What we're going to show you now with my help – with help from Andy Parsons and Mounir Ibrahim is how far the technology has come and what additional things it can do today.

With that, I'll turn it over to Andy Parsons for our first series of examples. (Applause.)

Andy Parsons: Thanks very much, Andrew, and thank you all for being here virtually and in person.

So I'm Andy Parsons. I oversee the Content Authenticity Initiative at Adobe and right up front I want to be clear about the relationship between the C2PA standards body, standards development organization, and the CAI at Adobe, which is a broad coalition of membership as well.

So I think of the provenance adoption curve as requiring layers in terms of what is possible. And they're not pure technology layers. This starts with a foundation in standards. And, as Andrew said, we're moving quickly with the 1.3 spec and the 1.4 spec after that to add the necessary support so news organizations, individuals, creators can take advantage of responsible use of AI and technologies that we ourselves at Adobe are constantly producing.

The C2PA is a small number of very committed individuals and companies. If I'm not mistaken, if you attend all of the C2PA standards development meetings you can spend upwards of 15 hours a week in meetings. And, astonishingly, some folks do, from Adobe, from Microsoft, from Truepic. And this is not because we have nothing better to do. We're all very busy, especially in the world of generative AI. It is a reflection of the urgency that we all feel around solving the problem with transparency.

The CAI has about 1,000 members. I think we just crossed the 1,000-member mark. It is run by Adobe, but it is a broad coalition of folks who are not writing technical standards, but rather signing up to endorse the idea of content provenance and adopting open-source that my team is working on

to reflect the innovations in the specification. So together they're highly complementary. It's quite a large and growing community of people who are focused on provenance. And we think we are well on the road to broad adoption. And today I'll talk about some examples of that adoption. By the way, the CAI include camera manufacturers, media companies of all kinds, and Adobe and its competitors as well, making creative tools for the next generation of creation.

This is where a dry run would have been helpful. OK. So the idea behind provenance, as Andrew said, is to be transparent about what something is because, effectively, as a baseline, if we don't know what something it, where it came from, or when appropriate who made it, we really don't have a good chance of understanding what our content is. So this is enabling good actors to behave responsibly. And over time, we expect that with ubiquity of provenance we will see that consumers learn to treat content that does not have provenance with a certain skepticism, as they should.

The word "provenance" comes from the art world. You would not buy a Pablo Picasso painting unless you were pretty sure that you could provably know what it is. And using cryptography and technology, knowing that math doesn't lie, as Andrew said, connecting provenance to media itself so that it travels wherever that media goes and is truly inextricably linked, we think is that foundational key. And that's what we're working on with our partners. The 1.3 spec also adds a number of generative AI affordances, one of which I'll talk about in a moment.

So I'm going to talk about three examples. You may have seen in the industry news that we, at Adobe, released something called Firefly just last week. This is Adobe's set of tools – generative AI models and tools – to provide sort of creative copilot for creative cloud users and others, so that generative AI can elevate the possibilities of creativity. However, generative AI models, coming from Adobe and so many others in this very fervent space around gen AI right now, can be used to do harm, right? If you can type in the name of a celebrity or a politician and make them do or say something that they never said or did, this could be extremely damaging.

With the 2024 election coming up, I need not tell this audience that there's some peril ahead. And again, without transparency, without a standard way to understand what something is, we will have a very difficult time telling trust from fact. So in the case of Adobe Firefly, released last week, we really put our money where our mouths are. And we did that by saying everything that comes out of Adobe Firefly will have a provenance mark, using the C2PA specification. So if you use this tool, which is in beta now, and its various invocations across the creative cloud as they come out, everything that emerges from these models will indicate what model was

used, what version of the model was used and, again, if appropriate, the identity of the person using the model.

That's evident in this beta version right now. We're committed to this idea of AI transparency. And there's even an affordance for artists to say: I don't want this material to be used for training. So if it's a photograph and it is not to be used to train other models, or even Adobe's model, we have a C2PA standard way to do both AI transparency and this notion of do not train. Now, technology's the easy part. Adoption and behavior change is the much harder part, which is why we're talking to all of you today. But these technologies exist. They're reified in the specification and in the examples that I'll talk about.

Sorry, I'm cricking my neck.

The second example that we have is – you know, kind of proves that government has a role to play. And in these early days, we do see some government offices adopting the technology. So we've been working with the Assembly Democratic Office of Communications and Outreach, the DCO, to enable content credentials in all of the photographs that they capture and publish. They're Photoshop users. They're customers of Adobe Creative Cloud.

And in Photoshop now, if you're a Photoshop user or you download the latest version, you have full access to what we call content credentials, which is the C2PA feature in Adobe Creative Cloud. And this is helping the DCO build more trust into an important source of content for citizens. All the images that they take and publish have this attached to them. And to be very specific, there is a little icon you can see on the right there. In this case, Evan Low. Photographs have been taken of Evan. An icon will appear on the website or wherever this image goes that supports content credentials and a consumer will be able to say, hey, not is this true or false is this really a picture of Evan, but seeing that little icon – it's an "I" with a little star around it – know that there's more information here. And over time, again, we expect that all content that you might see – video, audio, images, maybe even text someday – will have that icon or some indication that you tap or click and know that there's more information here. And if you want to know what something is, who's responsible for it, how it was made, that information should always be available to you.

So the DCO is pioneering here. But as Dana will speak about later, there's a role for all of you and all of government to play here to adopt this as well.

And the last example I'll share in this brief time we have together now is something that was published by the Rolling Stone I believe in December of last year. So the Rolling Stone undertook a pretty deep investigation into

the Bosnian war where some iconic photographs that you've probably seen by a photographer named Ron Haviv, who's a big support of – a big supporter of the CAI and C2PA, came under some scrutiny when the Rolling Stone tried to identify some parties that were responsible for some pretty egregious war crimes against the citizens. And the image dates back to the '90s. The investigation was not about content provenance, but it used content provenance as a key part of the investigation. And in fact, it used the Adobe CAI open-source to make some of this work. And the idea was simply: Can we prove that this photographer smuggled that photo out of Bosnia in the '90s, scanned it on a C2PA-compliant device so that there's cryptographic proof of what it is, who's responsible for it – not necessarily what's depicted, but the location and everything else we need to know about that image – so it can be used perhaps eventually in a war-crimes tribunal or criminal prosecution of some sort.

In this brief screenshot – and I'd encourage you to look at the article; if you look for Rolling Stone, Bosnia, C2PA you'll find all of this – it used decentralized blockchains as a complement to the C2PA idea and the C2PA and CAI open-source to display provenance on the Rolling Stone website and wherever that image ends up from here on out. We talked to the photographer. We talked to the makers of the equipment. We, happily, did not have much to do with the Rolling Stone's build of this website, and that's a testament to how straightforward the open-source piece is.

And again, that last layer on top of standards/open-source is user experience. This is one example of how the Rolling Stone took our open-source, built a UI on top of it that was appropriate for this particular investigation. And we expect, as Emily said, to see lots of different ways to display to consumers this notion of transparency. The key is to get started, to start working with open-source and with the C2PA, and start to give this to our citizens so that behavior change and adoption and sociological implications of this level of transparency can start to be better understood.

And with that, I'll turn it over to my colleagues to tell you about another example of the C2PA in action, Andrew and Mounir. (Applause.)

Mr. Jenks: Thank you, Andy. I'm back to talk about another example for C2PA usage. This one is related to an extension of what Project Starling had done with the historical determination of accuracy of images.

This is an application that is built to determine accuracy of current images. It marries a mobile application with the power of the Azure Cloud for translation and distribution to be able to take authentic images with location stamps in controversial areas – war-torn areas. The application itself is a simple implementation of the C2PA technology, but it shows some

of the power when it's applied to these use cases where this type of authenticity matters.

I'd like to invite Mounir up here to talk about the intention of this application and where it's being used today.

Mounir Ibrahim: Thank you, Andrew.

So when we built this application, we wanted to deploy this towards something urgent, something that's happening now. With a variety of conflicts around the world, we certainly saw the Ukraine example as an area that was semi-nonpermissive, meaning media couldn't access all these areas and information at time was ambiguous. Furthermore, we saw a variety of disinformation narratives coming from the Russian side to help muddy the waters.

With that in mind, we said: Let's identify a way and take this powerful technology we built, based on the C2PA open standard, and deploy it towards this area. But we wanted to be thoughtful in doing so, recognizing that we wanted it to be privacy first, secure these actors in high-threat areas. That led us to cultural heritage documentation. We thought, these are inanimate objects in known locations that represent a lower risk factor to deploy this as a pilot project. We reached out to U.S. government partners, like the State Department's Office of Global Criminal Justice, for thoughts, and then USAID.

We got introduced to Pact, an NGO, and its local implementing partner in Ukraine, the Anti-Corruption Headquarters. They were already risking their lives to document damage to cultural heritage throughout Ukraine. But they were doing so with standard images. And they were taking these standard images and putting them on a live interactive map that was incredibly impressive. We thought our approach with Project Provenance could directly support their efforts, and take it to a higher level to help reach their goals of advocacy, accountability, and awareness.

Why is this important? We recognize that these images are far more than just, you know, images for awareness and advocacy. They could potentially serve a purpose in accountability matters down the road. While we are not the experts on that area, our partners on the ground are firmly committed towards using these images to help future accountability and renumeration efforts in Ukraine. Furthermore, we've seen a surge of government support around the protection of cultural heritage around the world.

In fact, in 2017 the U.N. Security Council passed Resolution 2347, which condemns the destruction of cultural heritage, and which Russia supported and voted on. To date, UNESCO has at least 250 cultural heritage

institutions and landmarks that have been either targeted, destroyed, or damaged. Other estimates are up to about 2,000 different institutions and national landmarks. Our partners are working to document all of these using this technology.

The last point I'd make is perhaps even deeper than just immediate awareness or accountability measures. It's this concept known as the liars dividend. In my prior life I was with the State Department, and I worked intimately on the Syrian conflict. Very similar to the Ukraine conflict, we saw a surge of user-generated conflict coming out of that nonpermissive environment. Images and videos coming from everyday people's cellphones as they risked their lives to capture and share truth from the ground. These images were immediately undermined by bad-faith actors in places like the U.N. Security Council or in media and multilateral institutions by claiming that they were fake. The inability to authenticate and prove images and videos were in fact true undermined these people's causes.

That's what we are trying to help the Anti-Corruption Headquarters and our partners on the ground fight. Project Provenance is helping them prove the authenticity and transparency of this documentation. And that, to me, is one of the bigger goals of this project. I'll turn it back over to Andrew.

Mr. Jenks: Thank you, Mounir. And I want to call out that our colleagues from the Anti-Corruption Headquarters are actually on the line with us today.

So what was our objective in building this application? Our objective was to create something that allows people to document the material that they see in their everyday lives for later uses, without the ability for others, bad actors, to come back and say: This has been altered. This has been transformed. Because again, as Andy said, math doesn't lie. You can see from these images some of the original works that the Anti-Corruption Headquarters has already taken. These have been provided to news media organizations around the world so that they can authenticate the provenance of these images as well. And we continue to work with others to deploy these images where they are most useful.

You can see an example of the user interface that shows the verification of these photos, and provides the metadata that's attached to it. This is just one example of what provenance might look like from a small original project. With that, I'd like to thank you all very much for listening to this information about the various examples of media provenance in today's environment. And I'll turn it back over to Emily. Thank you very much. (Applause.)

Ms. Harding: All right. Some striking shots there, and some interesting ways to use technology.

I'm going to invite our panel up on the stage now. Suzanne Spaulding is going to moderate. And we have three guests that she will introduce.

Suzanne Spaulding: Thanks, Emily. And thanks to our presenting for presenting us such a clear picture of both the threat and a potential step forward for addressing that very serious threat.

We now have a terrific panel assembled to help us expand this conversation. As we think about the ways in which the threat that's been described here hastens this slide into a post-truth world, and what the implications of that are for democracy. And I do believe that this is a very clear objective of our adversaries, for example, who are engaged in information operations, to get us to a point where we give up on the idea of truth, of the idea of finding truth. And therefore, we lose the informed and engaged citizenry upon which a democracy depends.

So it is, I think, of vital importance that we think about ways in which we can arrest that slide into a post-truth world, in contexts that – where truth becomes so absolutely critical. I'm trained as a lawyer, and much of the work that I've done here at CSIS has been on looking at adversary information operations, really Russian information operations, targeting trust in our justice system. And you can imagine how deepfakes can impact the ability for a court, for a jury, to arrive at a conclusion, a consensus, between competing facts, if they no longer trust the things they hear, the things they see, right. And if we can't rely on the legitimacy of our justice system, what happens to our democracy, that fundamental pillar?

So these are really important issues. And we've got a terrific panel here today to discuss them. Ambassador Beth Van Schaack was sworn in as the State Department's sixth ambassador-at-large for criminal justice – for global criminal justice in March 2022. In this role, she advises the secretary of state and other department leadership on issues related to the prevention of and response to atrocity crimes, including war crimes, crimes against humanity, and genocide. Ambassador Van Schaack served as deputy to the ambassador-at-large in GCJ, the Global Crimes Justice, from 2012 to 2013. Prior to returning to public service in 2022, the ambassador was the Leah Kaplan visiting professor in human rights at Stanford Law School.

Dr. Matthew Turek assumed the role of deputy director for DARPA's information innovation office, or I2O – it's at DOD so it has to become an acronym, right – (laughter) – in May 2022. In this position, he provides technical leadership and works with program managers to envision, create,

and transition capabilities that ensure enduring information advantage for the United States. He joined DARPA in July 2018 as an I2O program manager and served as acting deputy director from June 2021 to October 2021.

Jessica Brandt. Jessica is the policy director for the Artificial Intelligence and Emerging Technology Initiative at the Brookings Institution. She's a fellow in the Foreign Policy Program's Strobe Talbott Center for Security, Strategy, and Technology. Her research interests and recent publications focus on foreign interference, disinformation, digital authoritarianism, and the implications of emerging technologies for democracies. She was the lead author on "Linking Values and Strategy: How Democracies Can Offset Autocratic Advances." She was previously head of policy and research for the Alliance for Securing Democracy and a senior fellow at the German Marshall Fund of the United States.

So a great panel. Let's get started. Again, let's start with talking a bit about the nature of the threat. Let's talk about the ways new technologies, and particularly AI has been – has been discussed, are changing the way that deep fakes not only are created, but also disseminated. Jessica, let's start with you.

Jessica Brandt: Yeah. Hi. Thank you. It's great to be here.

So it's advances in machine learning that make deep fakes possible, right? It's the "deep" in deep learning that gives deep fakes its name. So with enough computing power and data, it's now possible to create realistic-sounding audio clips, videos, and images using deep learning.

And I think there's probably two pathways where I think AI is sort of shaping the landscape. The first, as you gestured at, is, you know, it's enabling the creation of this content. And of course, we all know that it's – that sort of access to that capability is becoming more widespread. And then the second is that AI in the form of algorithms, you know, shape the way that content spreads around our information environment, and so they are impacting, you know, what is surface and what goes viral. So in these two ways, I think AI is sort of shaping our information landscape.

Ms. Spaulding: Matt, do you want to add anything to that?

Matthew Turk: Sure. I'll add a couple things.

Just sort of to emphasize that point of generative AI technologies I think is really the thing that is transformational here. Now, the ability to type a caption and get an image that represents that caption, that's a new,

compelling capability, and certainly one that could enable disinformation/misinformation.

If you look at the capabilities of something like GPT-4 and the ability to write software and start instantiating websites, now it's not just about the fake content but you can create, essentially, fake properties on the internet now to host some of that content, and you can do it quickly and at scale and with much less skill than you needed to do it previously. So I think those are some of the game-changing technologies.

Ms. Spaulding: So in a way it democratizes deep-fake creation and dissemination, I guess.

Ambassador, did you want to add?

Ambassador Beth Van Schaack: No, I'll just say I work generally in the justice and accountability space, and we've already seen how mis- and disinformation can infect that space, both from the prevention of atrocities but also in terms of responses in justice sectors. So much of the post-election violence in Kenya in the late 2007 and '08 was fomented by misinformation, and this was all created prior to the ability to create deep fakes, right? This was just stuff going viral that the election had been stolen, that there was violence everywhere, and that begatted additional violence. Now we have the ability to use synthetic media, generative AI, et cetera, and so you can imagine creating pieces of evidence that could actually infect a legal process.

So there was a new case before the International Criminal Court that was mentioned that involved images of a potential defendant committing a summary execution. And at best, if that had been a fake you would have had a lot of wasted time and energy upon the prosecutor's part to try and confirm whether or not that individual was, in fact, the individual depicted and should be indicted for that particular event. At worst, you could have had an amazing miscarriage of justice if that case had gone forward. And yet, we have an international but also a domestic judiciary that I think is quite new to this field, and so they don't necessarily have the ability to gauge whether or not they can rely upon digital imagery.

And it's increasingly being produced in justice processes around the world. Ukraine was mentioned, but even Syria was at one point the most documented crime base in human history primarily because of user-generated data. Now I think Ukraine has probably surpassed Syria. That's, obviously, nothing to be proud of or be happy about, but this is the world in which we're operating.

Ms. Spaulding: Yeah. I want to remind folks to – online. You've got a QR code. You can scan that QR code and enter your questions. I would encourage you to be thinking as our panelists are speaking here about questions that you might

have. And for those of you who are lucky enough to be here with us in the room, raise your hand and you'll be given a card to write out your questions so that we can offer those up to the panelists.

Ambassador, I want to – I do want to pick up on the point you were just making because I think we often tend to be very U.S.-centric. And you know, we're very aware here of the threat, as we see it, in our country and to our democracy. You do have a particular vantage point to be able to look globally. Are there – are there – and you've just described some of places you think it is perhaps most prevalent and certainly most concerning. Give us a sense of what that whole global picture looks like from the standpoint of these deep fakes.

Amb. Van Schaack: Well, now that we have global penetration of smartphones, essentially, everyone is a documentarian. We're not relying anymore on human-rights advocates or journalists, who may or may not be trusted to create verifiable and authentic content. Everyone is creating things. And they can manipulate those digital imagery and videos, sometimes because they want to highlight something so they'll add arrows or they'll zoom in, but sometimes they may want to try and manipulate those imagery in order to advance a particular narrative. And courts are increasingly having to deal with open-source information. It's never going to replace witness testimony, but we are getting to the point where it's taking up a large amount of the time and energy of investigators.

On the one hand, this is a good thing. We now have more evidence that investigators can draw upon in order to potentially identify responsible individuals. But we also have the opposite problem of the sheer volume of open-source information. And we need to be able to use machine learning in order to sift through that, to de-duplicate, to find the most authentic image, to find images whose metadata has not been stripped or altered in some way so that you can rely upon it, and that's where initiatives like the one we've just heard about are so critical. If they can lock down that metadata in order to be able to convince a judge or an investigator that wants to rely on that piece of information that this was, in fact, taken on this date at this place by this device by this human being at this time under these circumstances, then we can argue about what it depicts but at least we know that the image is what it purports to be.

And where we are lacking, I think, is in the fact that, you know, journalists are very good at open-source investigations; lawyers, not so much. You know, you may be better than me, but this is a new area for many lawyers. But the judges are even farther behind, I think. I mean, there are still judges who are dictating their judgments, right? And so we need to have a process whereby we can build trust in these images and also to teach judges how to evaluate these.

The Starling Lab – the Rolling Stone article, which I really recommend to everyone who's interested in this world, was contributed to by the Starling Lab, which I used to be affiliated with, at Stanford. I just had a briefing by the lab a couple of weeks ago when I was in Silicon Valley meeting with some tech companies on this sort of large problem set, and I can report that a domestic investigation has been reopened because of that investigation – because of the ability to identify the side of the individual who was just shown on the side by virtue of his insignia, where he was, other indicia of truth from that – from that digital imagery that was created, you know, years ago, before we could lock down all of this metadata. But that investigation has been reopened. And so this is the positive side of the new technologies that we're able to now utilize.

Ms. Spaulding: Yeah. That's great.

And you've talked a little bit about and we heard in some of the presentations the ways in which this is used – being used not only in a court of law, but as a weapon of war, right? Jessica, talk to us a little bit about the ways in which you see this not just currently, but as we go forward ways in which this can be –

Ms. Brandt: I mean, deception as part of war is nothing new, but I think what is new is, you know, the ability for virtually any actor now to create content, you know, that appears realistic that portrays an adversary's, you know, political leadership/military leadership taking action that they have not. And I mean, I can imagine a wide variety of sort of frightening scenarios about how this could be employed, whether it's to falsify orders to get troops to retreat or to stay, you know, or to take another action that, you know, was not intended by their leadership. You could imagine it being used kind of for the flipside, right? Not to get false orders to be followed but to get legitimate orders to be not followed by sort of incepting chaos in an adversary military or, you know, within society. You could imagine, you know, the creation of a justification or pretext for violence revealing, you know, plans for an invasion or something that needs to be sort of intercepted or preempted or any number of ways to kind of generate rifts within society, especially, you know, in contexts where there's sort of civil violence, civil conflict. And so I could go on but I think the, you know, the – there's a lot that's sort of framed in there.

Ms. Spaulding: Matthew, do you want to add anything to that?

Dr. Turek: I guess what I'd do is just open the aperture a little bit and it's not just military conflict, right? So there's certainly the ability for adversaries to more subtly influence population perceptions, drive narratives in competition phase, and we're essentially in competition worldwide all the

time, and so, you know, maybe the shape of conflict will change and there will be a lot more prepositioning during competition phase and a lot more dynamism around the narratives trying to influence the population, and whether that's through driving polarizing information or trying to create evidence for views or activities that never happened. I think those are also opportunities for adversaries to make use of generative AI.

Ms. Spaulding: Yeah. And I think there's a growing awareness – there's always been – in the nuclear context, there's always been an awareness of the potential threat from deception in that context in which things need to happen often so quickly, but I think there's a growing appreciation for the ways in which technology is seriously exacerbating that threat as well.

Ambassador, feel free to address that question about the context in war but also would love to hear from you about – more about the implications for human rights activists, for those who are trying to chronicle international crimes. I remember several years ago when I first came to CSIS and started working on disinformation getting a visit from someone who was working with the White Helmets and they had a huge challenge in this area.

Amb. Van Schaack: Yeah. No, it's a terrific point. And as Jessica said, the information space is a new domain of warfare, right? We've known that for a long time, but we're really seeing it now at scale, given all of these technologies and how dependent we all are on digital imagery, on digital technologies. In the human rights space, I think not just in the warfare space, we can also be concerned about using these imageries, et cetera, to foment violence.

We remember back in the Rwandan genocide in the 1990s the use of the radio, right, that that went and incited individuals, it identified individuals, it gave addresses. That was very much used to disseminate information. Now we can do that with everybody having a cell phone, and so we saw the Rohingya genocide and the role that Facebook played, obviously inadvertently but the platform was being used to foment violence there, and that very much magnified and augmented the ability to spread hate across the platform and to encourage communities to rise up against each other in sectarian violence. And so it's also a concern about atrocities prevention.

And your point about human rights advocates is really a key one. We saw from the earlier presentation the important role that these individuals play in apprising us in advance when there are risk factors, but they are also themselves at risk. And we want to make sure if they're out there creating content, they're doing so in a way that is safe, that is not drawing attention to themselves, and that is worth the effort because that content can then be used because it can be trusted and can be part of a chain of custody. And so I think authentication standards, like what were being discussed here, will

help very much so because judges and lawyers are used to the idea of a chain of custody; they want to be able to trace the murder weapon that was found at the scene that gets put in a bag, that gets put in an evidence locker, that gets brought to court, to be able to show each step along the way. We now have to do that digitally. That's what Starling was doing with their distributed block chain solution, but there's other ways that this can be done as well, in order to be able to say, nobody has touched this, this hasn't been changed since it was originally generated, and that, I think, is what's going to be important from the justice and accountability perspective so that we can use the materials that human rights advocates are creating.

But they're also at risk of being doxed and otherwise, and so we have to teach them how to protect themselves and how to disarm fakes that may be targeted against them, against witnesses, et cetera. The Kenya case that I mentioned, before the ICC, ultimately fell apart because there was rampant witness intimidation and tampering, and so the charges had to be dropped and the cases were essentially put on ice because they could not proceed without that witness testimony. And so if we can use digital platforms to create evidence, we can also use it to harass witnesses who are essential to justice processes.

Ms. Spaulding: Matt, Jessica, do you want to weigh in on –

Ms. Brandt: I'm reflecting on your remarks about the broader information competition and, you know, I think we've seen some, I think, remarkable moves by, you know, our government in the context of Ukraine, exposing and declassifying information to try to get ahead of, you know, sort of Putin's next moves, not to forestall the invasion – I don't think that was possible – but to try to shape the way it was perceived by publics around the world. And I just think of the very important role of open-source investigators in being able to corroborate government messaging. I don't know that the pronouncements of the IC would have carried as much weight if we didn't have, you know, a very vibrant and I would say much more mature than in 2016, for example, community of journalists and other researchers, and so I think these tools are also, you know, sort of useful in those hands because they can round out the picture and help our government compete in an information environment, drawing on democratic values, you know, doing so in ways that are, you know, reinforcing of the truth and rooted in truth. And so I think that's another important application.

Dr. Turek: I think that chain of provenance is particularly important perhaps in the legal context, diplomatic, other areas. You know, digital provenance techniques can help assure that, but it's going to take a while before they have sort of full penetration into the technical ecosystem. You really need to be able to, you know, control everything from the low-level processor on the cell phone to the operating system, the applications, the cloud, the social

media platforms on the far end of it. And so, you know, we really need a defense-in-depth strategy here where we leverage, you know, other techniques for finding associations between images, establishing provenance. Some of the technologies we've been developing at DARPA on programs like MediFor and SemaFor can really provide that defense in depth. In addition to places where we have that digital provenance chain, we can patch it using some of these other techniques. And of course, all of that, particularly to support legal cases, you know, needs to be established in the scientific community, support Daubert criteria, things like that, so that lawyers can take these techniques to court and use them to support their arguments.

Ms. Spaulding: Yeah, we need to have lots of innovation going on out there, right, and not think we've found an answer too prematurely.

Jessica, I really wanted to pick up on your discussion about leaning into the values, democratic values. You talk about truth but also transparency, right? One of the things that I think is so important about these efforts to be able to verify provenance is the extent to which they rely on transparency, which is our strength, right? I mean, democracies are much more comfortable operating in a transparent world. My colleagues have heard me time and time again talk about this concept of training to fight in the light. You know, if you train to fight in the dark, you could turn off the light and have the advantage. We need to continue to train to fight in the light because that's our advantage. And then we turn on lights so that our adversaries, who need dark corners to hide their corruption from their publics, for example, no longer have those – have that darkness because we've flipped on the lights, so it's a way of leaning into democracy's strength.

Ms. Brandt: Yeah, I would agree with that. I mean, I think a lot about the sort of emerging competition between democracies and authoritarian challengers, and I would describe that competition as chiefly asymmetric, right? I mean, I think open information environments confer tremendous advantages on democratic societies over the long run, but they create sort of legitimate and very real sort of short-term vulnerabilities that authoritarians exploit, and I think this is the paradox, right, where autocrats are actually quite vulnerable to open information and so this is – I think what we need to do, rather than kind of taking a tit-for-tat sort of narrow approach to responding to this new form of competition or this understanding that the information domain is a new domain of conflict, but rather do an audit of what are our strengths and what are their weaknesses and go on offense in the places that are most conducive to our success. So that's why I do think, you know, a persistent information campaign, sort of carrying the persistent engagement approach that we've applied to cyberspace and carry it over into the information domain but making sure that, you know,

our activities are grounded in, as you say, transparency, truth, and trusted information, and as I said, I think this can be, you know, an important tool.

Ms. Spaulding: Yeah. Great.

You all have touched a bit on the – really, what it comes down to is that we need to make it ubiquitous, in a way. It's not enough to just have a few key players using the defense in depth, the kinds of tools, particularly as, Ambassador, you talked about, the importance of ordinary individuals with cellphones, you know, documenting so much and being able to spark all kinds of activity, right? How do we teach those living in democracies about the importance of being able to establish that chain of custody, verify provenance? How do we get the technology world, where do we find the incentives to get things built in? And then how do we get ordinary people to care enough to turn it on, assuming that we've gotten it in the technology and it's – assuming it's, initially at least, not the default. How do we get people to care?

Amb. Van Schaack: It's a fundamental question. And I think fora like this are really important, to bring together these different segments. So those doing documentation need to be working with tech companies. I think app-based approaches are great because they can be immediately downloaded. So Truepic was mentioned. There's eyeWitness to Atrocities as well that has created an app that does the same thing, by locking down metadata and enabling stuff to be immediately uploaded, in a Wi-Fi environment, and then can be analyzed by lawyers. Getting the word out, essentially. I think governments can do that through our programming work.

I mean, we fund a lot in the civil society space, where we work with civil society actors who are often on the front line of documenting abuses and documenting war crimes. And so having – working with them in order for them to then magnify the impact amongst the communities that they represent in war spaces. We need to develop even some hardware. So there's a great organization called Videre Est Credere that creates these little clandestine cameras. Again, stuff can be immediately uploaded, so that people can wear them when corruption is happening or just to be observing the world around them as events are unfolding. These are all the types of solutions that we need to be working on. But then the key is, of course, dissemination, getting it out there.

You mentioned also – and I think you mentioned, Matthew – the idea of the Daubert Standard. Sort of when is expert testimony, and when are technologies acceptable within a legal process? Part of what we were trying to think about at Starling as well is how do we train expert testimony to be able to – experts to give testimony to be able to convince judges of the veracity and the reliability of these technologies and techniques.

There's a famous case of Patrick Ball, who was a statistician that works in this space, testifying before the Yugoslavia war crimes tribunal. And he's sort of halfway through his testimony, he had had a number of data sets that he was relying upon in order to identify who statistically was most responsible and who statistically was the victim community of an incident of mass violence.

And at one point the judge sort of threw up his hands and said: You know, we all went to law school because we don't like numbers. And his testimony was basically disregarded by the judge, even though it was looking at the highest statistical, you know, standards of proof. And so that's what we need to be able to overcome. And so being able to train experts to be able to explain these new technologies in essentially layman's terms so that it lays the foundation for the material to be able to be used in court.

Dr. Turek: I think platforms have an important role to play here. The question about how do we get society writ large to care I think is very challenging and very fraught. YouTube has done some interesting things where they have essentially trained viewers to ask questions and to think about disinformation. And I believe that has shown some effects in the – you know, for those particular viewers. But there are populations, like the documentarians, right, where they're going to care. It's their identity that needs to be protected, their reputation that needs to be protected.

You know, on our semantic forensics program, we've built the ability to defend senior leaders against deepfakes. It requires a significant amount of person-specific training data. But what if we could make that easier to use, make it quicker to build those models, democratize it, so that you could defend those individual documentarians? Again, that's a place where, you know, they're – I think the audience there, in that case the documentarians, would be interested and receptive to technology. And so that's a place perhaps where a technical offset could really make a – make a difference.

Ms. Brandt: Yeah. I think we're talking about two related but distinct challenges. One is, how do we build a society that's more resilient to disinformation efforts. And the other is how do we, you know, apply this kind of technology in the justice context. And in the first, you know, media literacy I think is an important component, but it's absolutely not going to be enough. I think it's wildly inadequate to ask, you know, an average citizen to go up against the well-resourced intelligence services of adversary countries, right? So if we're relying on, you know, my, you know, aunt in Tennessee, it's just not – that's not going to be sufficient.

So it's great to see the kind of partnerships between these various private sector actors, private sectors and civil society, civil society and government. There's a role, I think, for everybody to play in building a more resilient society. And there, I agree, I think the platforms have an important role in decreasing friction. You know, you talked about defaults, right? This is an important way of kind of increasing the – incenting good behavior. So that's sort of that problem set.

And then, you know, in the justice context, I think you're right. You're talking about actually targeting, you know, a very vibrant community, but a specific community. And I think there you can, you know, use the kind of levers that you're talking about to make sure that they, you know, understand what options are available to them, because I think they'd naturally be inclined to participate in ways that are conductive to improving transparency.

Amb. Van Schaack: And I would just add quickly, it's not about just creating technologies. I think we have to train people to use them. And so you can't just hand out clandestine cameras and say, good luck, off you go. You know, you have to teach people the kind of techniques that they can use, what to do if they're discovered, how to – you know, panic buttons so that your phone gets wiped. You know, whatever it takes in order to keep these individuals safe, who are taking great risks in order to inform the rest of us what is actually going on the ground.

Ms. Brandt: And there's ground to build on. I mean, I think there's been a recent – you know, a lot more attention's been paid to how investigative journalists and rights defenders can defend their cybersecurity, for example. There's all kinds of trainings for that. And so I just think there are communities of practice that exist. And so going to those communities of practice and making sure that this is, you know, understood, I think is important.

Ms. Spaulding: Yeah. I do think it's vitally important that as we work with human rights activists, for example, around the world, that we are realistic, and make sure they are realistic, about the degree, for example, to which their anonymity can be protected, the risks inherent in the use of this kind of technology, as you say, if they are, you know, taken into custody, et cetera. That's a real obligation that we have.

And on the public – raising public resilience, I mean, I do think this is a – this is a key issue, is how do we get to that broader public. And the question of how do we get them to care? Again, my colleagues will not be surprised that I'm going to take advantage of this moment to talk about civics education, because I really do think that a lot of this slide into the post-truth world comes from the slide down in trust in institutions and in democracy.

And that we can, over the long term in a sustainable way, begin to rebuild that trust by re-instilling a sense of what democracy is all about, how our government's supposed to work, and how individuals can hold institutions accountable and move us toward a more perfect union. So instilling a sense of civic responsibility is an essential part, I think, of making people care about whether they're forwarding something that's fake.

So how do we get people to want to look for that credential, or that certification, or whatever it might be at the end of the day? We've got to make the stigma of sharing something that is fake greater than the prestige of the first one to share. And I think that's a real challenge, but I think it is something that we've really got to do. And how should governments think about their role in trying to achieve scale in the kinds of technologies that we're talking about here? What is the role that government can play in, Matt, in this?

Dr. Turek: Yeah. I mean, I think there's a number of things that we can do. Certainly U.S. government produces media leveraging things like digital provenance for that media to help establish trust, to establish international norms, to work with international partners, and to create verification standards. It's certainly something that we can do. But again, that doesn't necessarily control the entirety of the media chain. So I think it's still going to be important to make investments in some of these defensive technologies, like the programs that we've been running at DARPA to detect, attribute, and characterize falsified media.

It would be great if there was commercial incentive to create those defensive technologies. I don't think long term we want the U.S. government to be the only source of funding or to really drive that industry. Perhaps we'll start seeing that inflection point relatively soon. I think that as the generative AI techniques become more broadly available, we see more sorts of attacks. I think there will be more opportunities for commercial industry to stand up. Again, I think that's a place where government can help.

Essentially all the technologies we've been developing at DARPA we publish regularly. Sometimes the source code and the software is available that can help bootstrap some of that commercial industry. We've been engaging with the research community to help offset just the – you know, the massive amount of investment that's been happening from commercial industry in generative AI techniques. Obviously, we're not going to match that one for one, but trying to find those places where we can provide some strategic leverage to create some of those defensive technologies. Again, I think there's an opportunity for U.S. government to act there.

Ms. Spaulding: Well, and we got a question from the audience that I think is directly relevant here. The authorities that the U.S. government has, certainly the comfort level that the U.S. government has in dealing with disinformation, for example, is really around foreign sources of disinformation. And yet, foreigners are not the only sources of disinformation. So, you know, how big a problem is that as we, you know, try to move forward and get these technologies accepted, and create a market for them? And there's a huge business cost, potentially, to deepfakes. I mean, I think businesses have already seen some of this. So I would be surprised if there isn't ultimately a commercial market for this.

Dr. Turek: Mmm hmm. Yeah, I mean, that's a really difficult question. The authorities – so let me just give you a little bit of perspective. So DARPA's a research organization. We actually have no operational authorities. We'll develop technology. We'll work closely with transition partners in U.S. government or with international partners to provide some of that technology. But broadly speaking, authorities are a really – are a real challenge U.S. government. And they're sliced up across agencies. And that's something that an adversary can actually take advantage of, regardless of where that adversary is located.

So, you know, that's something, I think, that, again, not a place for – there's not really a role for DARPA there, but I think that's something that needs to be examined. I think we need to look holistically at, you know, what are the authorities across U.S. government, who has them, and how does that match to what the threats look like.

Ms. Brandt: I mean, I think there's a question of what government can do, and then there's a question of what government should do. And I think a lot of the legitimacy, you know, especially in the contested information environment, comes from, you know, civil society being the source or the authenticator, you know, of information. So I just – I would also caution that there are limits, I think, to what government should be doing in the information space.

Amb. Van Schaack: I'll just say, as an office that has a tiny, comparatively – (laughs) – programming budget, we have built into all of our notice of funding opportunities, that are generally focused on civil society, that's where my office tends to fund as many of the Human Rights Bureau, and AID, and others also do as well. We're encouraging civil society actors to build consortia that includes some sort of tech component, so that they have access to the most sort of cutting-edge thinking as to what's being done in this space, and they can integrate that into whatever they're producing – whether it's documentation, whether it's training, whether it's capacity building of justice actors, et cetera.

The second thing I think government can do is use its convening power. So just this week in fact members of my team who focus on tech are in Geneva. They brought together a meeting of tech companies and the various investigative mechanisms, devoted to Syria, to Myanmar, the ICC is there, others. And they are having a series of meetings talking about these issues – the intersection of tech and the justice sector, including all of these questions of authenticity and provenance. And so that's, I think, another contribution that we can make.

And then we do have, of course, criminal and sanctions authorities that can be used for individuals who will – who would try and peddle in mis- and disinformation. And of course, we have our own First Amendment protections. And so finding the line there is, I think, a critical one. The mention of parity was made, et cetera. So getting that right, of course, is critically important, but those authorities are available to use to identify and to impose some measure of accountability on those malign actors.

Ms. Spaulding: Yeah. And that means resourcing those law enforcement investigators and prosecutors to do that if we think it's a serious threat, right?

So the convening in Geneva sounds interesting. That's a U.S.-led convening?

Amb. Van Schaack: It's at the embassy.

Ms. Spaulding: Yeah. So what countries – as we look around the world, what countries and multilateral or international organizations are tackling this? Where can we look for some lessons learned? Are we – are we really out in front on this, or are there others who we can look to?

Amb. Van Schaack: I think – I think there's a real community now. I mean, thinking about the Estonians, right, I mean, they were – they were the subject of one of the first cyberattacks in history so they are, you know, way ahead of us, I think, on the defensive side, certainly. But I think your point is a good one, which is this has to be multilateral. This cannot be a single government doing this. We have to start to create global standards and global expectations, and then ultimately a global response to be able to address what is a global problem that doesn't respect international boundaries at all. And so we need to be able to respond appropriately.

And I think this is, again, a role that government can play, which is to sort of bring people together to be able to debate and discuss. We have an Atrocities Prevention Working Group that meets every other year and tech has always been a component of that conversation. How can we as governments that are a part of this working group empower and be on the lookout for and be aware of what's being done in the tech space that is

going to impact the ability to prevent atrocities and to respond to them after the fact?

Ms. Spaulding: Great.

So looking at the range of technological solutions to this very serious threat, Matt, talk to us a little bit about how secure they are. We've got a question from the audience asking whether these provenance-assurance technologies will be implemented with quantum-resistant encryption, for example.

Dr. Turek: Yeah. So interesting question, and I guess I would defer the specifics of the implementation to the folks that are – that are actually working on it.

What I will say broadly is, again, I'll argue for a defense-in-depth strategy. You know, coming from an office at DARPA that also does secure and resilient systems and does a lot of work in cyber, if you have enough skill and enough resources oftentimes you can overcome sophisticated defenses, but they provide a significant barrier oftentimes, right? So, you know, it's one thing for a sophisticated nation-state to have the ability to alter stuff; you know, a single individual working alone, you know, be nice to be able to take them off the table, take low-resource organizations off the table. So one is just that notion of, you know, what's the – what's the barrier we're imposing.

That defense in depth allows you to impose multiple barriers. You know, that's one of the technical strategies that we always adopted on our media forensics, our semantic forensics programs, is that there isn't just a single detector. We're looking for multiple kinds of evidence – you know, digital, physical, semantic evidence. Sometimes there's tens of detectors. You have to fool all of them in order to really accomplish a sophisticated media manipulation. Adding that digital provenance chain, again, another layer to those – to those defenses. So, you know, I think it's not so much a binary question of, you know, can those defenses be overcome, and I think it's really a question of what sort of adversaries, what sort of skill and resources does it take to overcome those.

Ms. Brandt: I don't think there's any single technology that is going to solve this problem for us. And I think, you know, as you gestured at, it's the liar's dividend, right? Once we live in a world where we can't trust what's in front of our eyes, you know, that is – that's an enormous challenge for a democratic society, you know, because those who are willing to can just say – you know, sort of sleight of hand – "it's not me." There was even, you know, I think already a case of a – of a(n) AI-generated, AI-enabled, you know, search engine hallucinating and, you know, making a claim about a piece of content that never existed, right? And so – and then at least all

kinds of challenges of trust when you can – you know, you kind of say, well, why did you take this off the internet? I never did. Well, the search says it, right? So I mean – and search is the place – you know, lots of studies have shown, like, search is the place that people – the most trusted, actually, place on the internet.

So I think we're in for a ride. And there – you know, these are important and beneficial technologies, but they won't, you know, solve the problem on their own.

Dr. Turek: Well, and it's certainly, I think, important to say that a technical solution is not going to be sufficient by itself, right? You need civics. You need a societal perspective, the willingness to pay attention and to engage with the media. It's not just we're going to come up with a silver-bullet technical solution and we'll all be safe.

Ms. Spaulding: Yeah. Yeah. And borrowing from the – we just did a big program on resilience in the context of climate, cyber, supply chain, and workforce, but borrowing some of those concepts about assume that – in the cybersecurity context, which is the one I know best, you know, assume that you've been hacked and plan accordingly. Assume that you're going to be hacked and plan accordingly. It, you know, sort of gets to this defense in depth, but you know, the idea that you might not be able to prevent a sophisticated actor from spoofing, say, your credential, but you could – you can build in a way to detect that, right? Detection becomes critically important. And think about all of the other ways that you would be able to build in trust.

Ambassador, you, I think in your work at Stanford, mentioned the work that you've done on blockchain. What's the role of blockchain in a similar way, demanding an adversary make the exact same change, you know, in so many places?

Amb. Van Schaack: Yeah. What Starling was focused on was holding digital data in a distributed way in order to ensure that it doesn't – hasn't been tampered with. The idea was kind of a chain-of-custody type of approach that's essential to authentication, and then coming up with ways in order to explain that and to get it integrated into without setting the bar too high. I think that was also a challenge that we were grappling with, which – we can't expect there to be, you know, blockchain everywhere, right? We have to be able to do it at lesser levels. And so, again, this gets to this point of being able to have interventions that are appropriate at different levels of technological sophistication. We can't expect everybody to have amazing cellphones that can do all of this incredible encryption, right? We have to be able to survive at a – at more of a low-tech, although maybe not an analog level, but a lower-tech analog – a lower-tech level.

But I think universities play a really key role here, and research institutes, to be able to do the sort of deep thinking about not only what technology is available, but what are the consequences of that technology. And we know the sort of, you know, Casio watch phenomenon, that something that used to be, you know, only in the hands of, like, you know, the purest of technologists will suddenly become available everywhere over time. And we're already seeing that with respect to, you know, smartphones, for example, even in very – you know, I was in Cox's Bazar recently, which is the world's biggest refugee camp, and everybody had a phone. Everybody had a phone. They barely have a roof over their head, but they have a phone. So that's the world we're living in.

Ms. Spaulding: Well, I think we have to – part of the education of the public and ourselves is a recognition that technology is rapidly outpacing our ability to even conceive of the ways in which it is going to change/has the potential to change our world for good and for bad. We're all getting a taste of that with this, you know, generative AI that is – that has gotten, I think, out ahead of our ability to anticipate. And we see the same phenomenon here.

But wonderful to hear that there are smart, innovative, hardworking people who are trying to get ahead in terms of developing ways for us to at least have a – make this a fair fight as we fight for a world in which truth does exist and can't be so easily dismissed. And so I want to thank the three of you for a great discussion and for what you do each and every day because I think it is absolutely fundamental to the defense of democracy. So thank you very much. Yeah. (Applause.)

Ms. Harding: All right. Thank you so much to our panelists, and to Suzanne for an excellent job moderating as usual.

Last but not least we have Dana Rao, who leads Adobe's legal, security, and policy organization. He's the executive vice president, the general counsel, and the chief trust officer. So who better to wrap us up today with a discussion about this combination of technology, trust, and democracy? Prior to joining Adobe, he was with Microsoft for 11 years as the general counsel for IP and Licensing, so I think he's got this issue covered from all angles.

Welcome, Dana. (Applause.)

Dana Rao: Thanks, everyone. I really appreciate the time here, and it has been a great opportunity for me to get back to D.C. I went to GW Law School. I met my wife there, and I haven't been here in 25 years with her, and we got to spend some time with the cherry blossoms yesterday, which was quite nice. When we were here, it was pre-Instagram so there's a lot more posing

going on now – (laughter). I'm not actually sure that anyone cares that there's trees, but at least they're outside.

It's really great to have the conversation we're having about the issues of misinformation and AI, and when I think about the problem, when I think about how to think about the problem, I think about it as seeing is believing, right? That's the problem. We, as people, tend to believe what we see and hear, right? We have an emotional resonance to things that we watch, that our brains are wired to believe things we see and hear, even more so than the written word, right? Before, 10 years ago, you got that email from the Nigerian prince about the $10 million you needed to pay, and you just paid it, right, and you just wrote that check.

But you've been trained, right, to be a little more skeptical of that written word, and I think people are a little more skeptical of what they read. But the images, and audio, and video have an emotional resonance, and it's harder to get people to not believe what they see and hear. And that's the key that we're talking about here.

And you can see the multi-faceted problem in all kinds of ways – you've heard a few of them – but, you know, you can imagine right now – and I'm saying you can imagine, but these are all actual examples – your CEO send a video or audio recording to the finance department, says, transfer some money over to our customer account. How do you know that it's your CEO sending that, right? You don't know.

Government leaders, as we've heard, right, can – we saw last week there was a fake audio of President Biden expressing concern about banks, right, and in lieu of the SVB crisis, right? That can cause an exacerbation of a bank run if you believe that audio is real.

And then, you know, there were images last week of Donald Trump being arrested prematurely, and again, starts little violent protests. Now a pro tip from Adobe Photoshop: three legs on a human, rare for that to be true – (laughter) – but the – as the – as the AI gets better, you're going to – we're going to have two legs.

The other pro tip right now that all the other AI-generated technologies are struggling with – we launched ours last week – often there's six fingers in the hands. Just pay attention to the hands. Sometimes they – I'm not sure why the AI is struggling with the number of fingers in a hand, but that is one of the things when you want to look at such deep fakes, look for the number of fingers on the hands.

And so we were in India a couple of weeks ago – my CEO and I – and he met with Prime Minister Modi, and he's talking about the same problem, right?

There's been fake videos that have caused protests in India, violence, and it's just as urgent a problem for India to solve, for every government to solve, every democracy to solve, as it is here in the United States.

And so you think about that span, you're like, wow, this is everywhere, right? CEOs, it's finance, it's government. It's also you guys, right? Every human here, right, when you think about that Nigerian prince email that I just mentioned, and you think about all the posing that's going in at the cherry blossoms, AI can train off of any image, any video that you are uploading right now, and it's pretty easy to simulate people's images in video and audio with a little bit of data on these models. So imagine your elderly parent getting an email from you with your voice on it, your video, asking you to transfer some money over to – them to transfer money over to your bank account, right, or please hit this link and type in your bank account information, right?

So this isn't just a government problem. This isn't just a finance business problem. This is a problem for everybody to think about how are we going to solve it, and it's really important.

What the good news is, is that, unlike probably 90 percent of things like this where you hear briefings about terrifying problems, we actually have a solution, right? We've been working on this solution collectively for four years because we saw the problem of AI coming, and the deep fakes coming, and the soon-to-be-real possibility you're not going to be able to tell the different between fact and fiction online.

So we've got a solution, which is really great. We've heard Andy talk about it. But in the coalition, the open standard we've made Adobe's implementation of that is called content credentials. It's in Photoshop right now. You take a – you take an image, you put it into Photoshop. You say: I want content credentials on. Saves your identity, where you took it, when you took it, and edits you're making to that image. Goes with it wherever another CAI member goes who supports it. Wall Street Journal, AEP, Reuters, these are all members. You see the image there, you can see what happened to it. You can see that credential, right? People are going to be able to see that transcript of information, decide for themselves whether or not to trust it.

Because the problem is twofold, right? It's not just everyone's getting deceived. It's right after everyone has been deceived, they're not going to believe anything they see or hear online anymore. And that's the real critical problem. That's the problem we were worried about four years ago. Once democracies don't have a way to communicate the truth, how is anyone going to be able to do anything anymore? Is that – was there a wildfire? Was there a war? Was there a protest? Is there an emergency? Is

there not an emergency? You're just not going to believe anything. So it is critical to be able to prove authenticity in order to do anything.

So when you get – when your parent gets an email from you and there's a video, they should see a symbol on that video to know that it was really you who sent it. When that poor finance person working in the bowels of the organization gets that voicemail from their CEO, there better be authenticity with it before they decide to transfer money. And when you see a message from President Biden, there better be an authenticity symbol so you can go verify for yourself that it was him. And if you don't see it, then you shouldn't believe it. And that's the digital literacy issue. That's the public education issue that we've been talking about so much, right?

First step, get the technology out there. We're really excited, not only do we have it out there for imaging, the specification that we've created – the open specification we have created has video and audio in it. And the leading image and audio companies are also part of this coalition. So you have a specification where you can start building provenance right now everywhere. And it's exciting that we've gone from, you know, one company four years ago to over 900 members of this coalition right now. So everyone understands the problem. Everyone understands the solution. They're building technologies. That's the exciting part.

The next step is implementation. I mean, you've heard people talk about this already today, right? I'm excited that CSIS is going to be showing images of this event with content credentials. That's great. Andy mentioned there's one government office in California that took it upon themselves to say: I'm going to start communicating with my citizens with content credentials. And then they forced that implementation to happen. Every single person in this room works for an organization that could be publishing their content through CAI. Everybody in your individual capacity, in your capacity as a government official, you all could do that.

The way we're going to drive implementation is everybody saying: I want this implemented. I want to be able to speak authentically. I need technology to let me do this. The technology's there. The implementation is easy. We actually open sourced our own code to made it even easier for everyone to build this. So everybody listening, here in the room, I know there's over 150 people online, you all can help with this solution. Just like CSIS has decided to do content credentials for this event, all of your organizations can do the same thing. That's the next step, everybody getting together and saying: We want this. We want our organizations to do this. This is critical. And I need to be able to speak authentically.

Now, the government has an amazingly important role to play. In Europe, they're passing the AI act. And they're looking hard at putting provenance

as a requirement for media that gets transmitted from the social media platforms. India is looking at the same thing. United States could also be looking at the same thing. What are the requirements that we're going to have for the way media gets shared online? Are we going to put in a requirement for transparency? I think we should. We're a democracy. Every democracy should.

Also, digital literacy, critical. We've talked about this. It's important for people to know that this tool is out there when it's out there. People are going to need to know, just like we talked about with phishing. We've done a bunch of education around that, but you can also think about it as a nutrition label, right? This is the – this is what's in this content. Click on this, find out what happened to this content, right? That's what you should be looking at when you think about this icon. Before you consume it, know what you're consuming, right? We can provide that education. That's the role of the government and they can do it.

So a lot has already been said, so I don't want to go on and on and on and to recap everything – what people have talked about, the strategies that we have to take on this problem, but I am excited that we have the technology here. When you see AI-generated work coming, as we've mentioned, there's so many new technologies – and Adobe launched its own last week with text prompt, where you type in, you know, cat with a hat on a deck on a boat, you get the image – a new image entirely created – it's another place where you can think about where – how is content credentials, how is authenticity going to play. Adobe's coming out in that tool, is automatically turning on content credentials to say this was an AI-generated work. So every time you create something with our AI-generated tool, it's going to say it was AI-generated so people know. It's another place where you're going to see content credentials be really important.

And in that case, it's not quite the misinformation case. It's absolutely a misinformation case, but there's also a market out there for people who want to know whether AI created something or whether humans created something. We have creative professionals who are very interested in being able to distinguish themselves from robots, and that's going to be a new – a new interest in transparency.

So you can see the cases are going to keep multiplying. The value of authenticity is going to keep multiplying. When you see where AI is going, where audio is going, where media is going, the cases – the use case for authenticity is growing. And I'm excited we have this solution. I'm excited we have the people here who are already thinking about the solution and how to develop – how to proliferate it. And I'm really confident we're going to get there. So thank you for your time. (Applause.)

Ms. Harding: All right. We have reached the end of our festivities. I want to thank everyone for coming out this morning for some pastries and some conversation. I know it's a Monday morning and I hope it was a great way to start it for you, a little bit of inspiration in the face of some really tremendous challenges.

We heard about the difficult challenges of protecting democracy in an era where actors seek to make nothing true and everything possible. We heard about potential solutions, from content provenance to civic education. I want to say a big thank you to all of us who – all of you who joined us online.

And also, thanks to Andy, Andrew, and Mounir for their hard work as the C2PA team. I know they are donating a lot of time they do not have to this initiative, and it's greatly appreciated. Also, thanks to Beth and Jessica and Matt, our wonderful panelists; to Suzanne, our moderator; and then, finally, to Dana Rao for coming out today and talking about the wide scope of this particular problem. Thank you so much. (Applause.)

(END)