Testimony of Mounir Ibrahim, Executive Vice President, Truepic

Before the U.S. House Committee on Oversight and Accountability

Subcommittee on Cybersecurity, Information Technology, and Government Innovation

Advances in Deepfake Technology

November 8, 2023

Chairwoman Mace, Ranking Member Connolly and Members of the Subcommittee thank you for this opportunity to brief this committee.

My name is Mounir Ibrahim, Executive Vice President of Truepic, a technology company focused on transparency and authenticity in digital content. We create what we refer to as the "infrastructure of authenticity" necessary today for business, government and society.

Prior to my time with Truepic, I was a Foreign Service Officer with the U.S. Department of State. Serving as a Foreign Service Officer was one of the greatest honors of my life and my time as a diplomat in Damascus, Washington, Bogotá, Istanbul, and at the United Nations in New York directly led me to advocacy for transparency and authenticity in digital content and my work today.

While these two worlds may appear quite different, let me explain the connection. My first posting with the Department of State was at the US Embassy in Damascus in 2010, about six months prior to the start of the Arab Spring. As a political officer, I was tracking issues related to civil society, opposition, and religious freedom prior to and during the start of the Arab

Spring. I witnessed the first several months of anti-government protests throughout Syria in person. I saw protestors risk their lives, being beaten and attacked in front of me as they rose up against the Asad regime. I also saw many more people risking their lives to document the violence with their smartphones. Many times, those documenting the violence on devices and smartphones were targeted before those actually protesting. It was an eye-opening experience that highlighted the power of digital content and audiovisual evidence.

Later, as an advisor to two different US Permanent Representatives to the United Nations I saw imagery from conflict zones regularly enter into the debate at the UN Security Council - the world's foremost forum for debate on international peace and stability. There, I saw images of chemical weapons use, attacks on schools or hospitals, or violence against civilians regularly get undermined by claims that they were not real[1]. All of it was regularly and easily questioned for being potentially fake or edited.

It was a highly effective strategy used by countries and critics, who wanted to undermine reality. Images and videos people had given their lives to record were easily dismissed. The *pretext* that digital content *could* be fake or edited beyond recognition created an opening for bad actors to undermine all of user generated content. As the volume and velocity of images and videos coming out of Syria increased, this claim - that user generated content could not be trusted because there was no way to know for certain that it was authentic - became even more common. Keep in mind this was **before** deepfakes or Generative AI even existed.

Today, this strategy for undermining reality is now commonly referred to as the "Liar's Dividend.[2]" Without widely adopted, interoperable standards for transparency and authenticity in digital content, bad actors benefit from the rapid increase in fake and manipulated imagery. It makes their false claims that a real image or video is fake more believable, giving them the ability to sow doubt in what we see and hear online.

---

[1] "Syria Says Photos Alleging Mass Torture Are 'Fake" Time Magazine, Jan 22, 2014, accessed Nov 3, 2023 https://world.time.com/2014/01/22/syria-says-photos-alleging-mass-torture-are-fake/

[2] Chesney, Robert & Cintron, Danielle, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security:* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954

In my opinion, this is one of the greatest challenges we face today as synthetic and AI-generated content proliferates at a rapid rate. Some estimates are that in 1-2 years, 90% of new digital content created online will be wholly or partially synthetic.[3] Without wide adoption of interoperable standards to clearly differentiate authentic content, AI-assisted and fully generated content, our entire informational ecosystem will be at risk.

I would like to emphasize that while my experience prior to Truepic is related to geo-political, humanitarian, and societal crises, the same challenges exist for every business and even every person on earth. Our shared economies and industries will also be at risk.

Simply put, we have digitized our entire existence and whether you are a government, business, or an individual - we all rely heavily on what we see and hear online to make decisions every day. This includes personal decisions like who we date, who we hire, what we buy, what we rent, who we vote for and, ultimately, what we believe. It also includes business decisions like: does an insurer pay that insurance claim based on pictures? Does a bank execute that loan based on a digital application? Can you virtually audit a supply chain, Is that vendor or customer real? And many more.

If we do not implement a transparent ecosystem, to tell the difference between authentic, manipulated, and synthetic content, our economy, society, and democracy will be under serious threat.

C2PA: An Interoperable Standard for Scaling Authenticity

This is a really challenging problem. There is no silver bullet to solve image and visual deception, no panacea that immediately tells you real from fake at scale, in real time and perfectly. Though many continue to work on detection mechanisms, we do not believe or see a

[3] "90% of Online Content Could Be 'Generated by AI by 2025,' Expert Says," *Yahoo Finance*, last modified January 13, 2023, accessed November 3, 2023, https://finance.yahoo.com/news/90-of-online-content-could-be-generated-by-ai-by-2025-expert-says-201023872.html.

future in which one mechanism can detect synthetic from authentic, edited from manipulated and verify associated metadata in real time or scale.

That is why we believe there needs to be a transparent ecosystem. If we cannot tell what is fake, can we create transparency around what is real, edited, or synthetic so that people know it when they see it?

Creating a transparent ecosystem for digital content and information, cannot and should not be done by any one entity. It is a large effort that will require years of work, effort, and education from technologists, civil society, and government.

Despite this herculean task, I am pleased to say that a lot of work toward a more transparent ecosystem is already taking place, starting with the Coalition for Content Provenance and Authenticity (C2PA). Founded in 2021, C2PA is a coalition for authenticity in digital content supported by Truepic, Adobe, Microsoft, Sony, Intel, the BBC, Publicis Groupe and many others, like WITNESS who is here with us today. The C2PA developed the world's first open standard for digital content provenance.

The basic concept of provenance is attaching the facts or history of a piece of digital content directly to the file itself. These facts may include the system that created the content, its date, time, location, and additional information based on user preferences. The C2PA standard is structured in a way that makes it interoperable - meaning information aligned to this standard will be able to flow across the internet to any compliant platform, software, phone or device while maintaining the transparent information. As the media file undergoes changes along the way, the chain of provenance adds in this new relevant information, compiling a more complete history. The C2PA is also designed to be tamper evident, meaning that if the chain of provenance is broken, users will be able to see that an unknown change of some kind has been made to the file.

The purpose of C2PA is for content consumers - those viewing images, videos, or any other content online - to have the option to view a piece of content's history transparently before making a decision of consequence.

The C2PA standard can be applied to any kind of digital content - from authentic to synthetic, or anywhere in between. It also works with multiple file formats like image, video, audio and others. It simply facilitates exposing both the origin & history of content to an end consumer.

Imagine if every time a Gen AI platform produced a synthetic image or video it was automatically signed with the C2PA cryptographic hash at the moment of creation. Then, when a content consumer sees this image or video they would immediately have a visual indicator - known as Content Credentials[4] - visible to inform them this content was created with Gen AI.

For an authentic image captured, Content Credentials are also available but with privacy top of mind. One notable difference is that for an authentic image, the photographer would have the option to add Content Credentials if they want to. It would not be the default and is completely optional for any user. The C2PA spec also allows for the redactability of certain fields, like location, that users may not want to share.

Truepic's Approach:

Truepic supports the C2PA open standard because we believe *interoperability* is an essential part of robust authenticity infrastructure. Digital tools, systems, and platforms need to be able to speak the same "authenticity language" so that essential information, for example whether something is AI-generated or not, is not lost, scrapped or removed as digital content travels between different environments.

We develop a suite of products that empower businesses, organizations, systems, and cameras to create authentic and transparent digital content at internet scale.

[4] ContentCredentials.org: https://contentcredentials.org/ Accessed November 3, 2023

Our technology and products boil down to two main areas: first, helping secure what is authentic, and second, adding transparency to generative media .

Authentic Imagery - Secure Capture Camera Technology:

Since 2015, our core thesis was that we would build camera software so that when someone used it, the viewers would know that the resulting image or video was authentic. We built what is referred to as Secure Capture Camera technology. This technology authenticates data and pixels from the moment of capture. The camera uses the C2PA open standard to embed that information cryptographically into the media file so that it is preserved and can move to compliant systems. Our camera technology also detects if the user is taking a 'picture of a picture' to ensure that the image or video captured is actually live. To date, our secure camera has been used in over 150 countries around the world and relied upon by hundreds of the world's largest businesses ranging from Equifax to Ford Motor company. Businesses rely on user generated content for verifying insurance claims, bank loans, "know your customer," and more.

I would also like to highlight our work with Microsoft to deploy this type of authenticating camera in Ukraine through a program called "Project Providence.[5]" Because both Truepic's camera and Microsoft's cloud align to the C2PA open standard, our technologies are interoperable. Over the course of this year, our camera technology has been used by a USAID partner to document the destruction of nearly 600 cultural heritage sites in Ukraine. Prosecutors in Ukraine have since opened at least 10 different legal investigations using the authenticated images from this secure camera.

Allow me to explain how the same approach can also be applied to AI-generated and synthetic media.

Transparency in Gen AI - Cryptographic Signing Technology for Disclosing AI Elements:
Truepic expanded its work on image authenticity to encompass transparency in AI-generated content. Building on top of the C2PA open standard, Truepic's technology cryptographically

[5] Project Providence: https://www.projectprovidence.io/ Accessed Nov 3, 2023

signs AI-generated or AI-altered media at the moment that it is created or edited so that these AI elements can be recognized by and disclosed to the viewer in the form of Content Credentials. We believe that giving viewers this information about whether pixels were AI-generated or not is an essential protection for American consumers, businesses, and society to mitigate certain risks.

Generative AI and synthetic media have become widely accessible and highly sophisticated over the last year and half. There is growing consensus that disclosure about whether something has been AI-generated or altered is an essential protection for anyone making decisions based on digital content.

Alongside 18 other industry leaders, Truepic is a proud supporter of the Partnership on AI's Responsible Practices Framework for Synthetic Media which specifies disclosure as a core best practice for the creation and distribution of synthetic media, and lists C2PA Content Credentials as a leading disclosure mechanism.[6] By capturing how a piece of media was created, in the moment that it is created, and signing that information into the media file itself, C2PA provides a robust technical foundation for this kind of essential disclosure.

In April of this year, we worked with the creative team at Revel.ai and author Nina Schick to illustrate what Content Credentials look like in a synthetic video, releasing the world's first transparent "deepfake" video[7] to highlight the possibility of this technology.

In October, we launched a space in partnership with Hugging Face that allows anyone to use C2PA Content Credentials to disclose that their creations are AI-Generated. This was a notable step to democratize transparency in AI-generated content to open source models[8].

[6] "PAI's Responsible Practices for Synthetic Media," *Partnership on AI - Synthetic Media*, accessed November 3, 2023, https://syntheticmedia.partnershiponai.org/.
[7] "Mirror of Reflection" https://www.youtube.com/watch?v=NhvkG8G4PN8 Accessed November 3, 2023
[8] Hurst, Alicia "Making AI-Generated Content Easier to Identify," Hugging Face Blog https://huggingface.co/blog/alicia-truepic/identify-ai-generated-content October 5, 2023, Accessed November 3, 2023

Most recently, in partnership with Qualcomm, we have enabled C2PA content transparency and authenticity at the chip level in their new Snapdragon chipsets.[9] As generative AI now moves on to devices, we believe that baking in transparency at the chip level is essential. We believe this is a watershed breakthrough that will enable any smartphone to deliver transparency right at the moment of creation.

It is worth noting that Truepic is not alone in applying Content Credentials and the C2PA open standard to Generative AI systems. Adobe launched the same capability in Adobe Firefly[10] and its creative suite.[11] Microsoft's Bing Image Creator also applies C2PA to outputs from Open AI's DALLE.[12] Stability AI has also announced implementation of Content Credentials in their AI API.[13]

How Government Can Help Mitigate Challenges

If possible, I would like to offer some thoughts on how government can help to mitigate the risks of AI-generated and synthetic media by leveraging the ongoing work on transparency and authenticity to protect Americans and advance national goals.

First, various agencies and facets of government, including this Subcommittee, have an incredibly unique platform and ability to educate and raise awareness. Events just like this hearing are critical to raise awareness and hear from the wide range of stakeholders who are working on this problem. I would urge this committee to continue holding forums and hearings like this and speak to a wide range of stakeholders.

---

[9] "Truepic Unveils Watershed Gen-AI Transparency Directly on Devices Powered by Snapdragon Mobile Platform," October 24, 2023, Globewnewswire https://www.globenewswire.com/news-release/2023/10/24/2765978/0/en/Truepic-Unveils-Watershed-Gen-AI-Transparency-Directly-on-Devices-Powered-by-Snapdragon-Mobile-Platform.html

[10] "Content Credentials for Assets Generated with Adobe Firefly," accessed November 3, 2023, https://helpx.adobe.com/content/help/en/firefly/using/content-credentials.html.

[11] "Learn about Content Credentials in Photoshop," accessed November 3, 2023, https://helpx.adobe.com/content/help/en/photoshop/using/content-credentials.html.

[12] Kyle Wiggers, "Microsoft Pledges to Watermark AI-Generated Images and Videos," *TechCrunch*, May 23, 2023, accessed November 3, 2023, https://techcrunch.com/2023/05/23/microsoft-pledges-to-watermark-ai-generated-images-and-videos/.

[13] "Stability AI Previews Enhanced Image Offerings: APIs for Business & New Product Features," *Stability AI*, accessed November 3, 2023, https://stability.ai/blog/stability-ai-enhanced-image-apis-for-business-features.

Second, I do not think we can just legislate our way out of this problem, but there could be specific legislation and government action that is helpful. The recent Executive Order section 4.5 specifically addresses concepts related to transparency in digital content and instructs agencies to begin examination, standards, and even piloting/testing.[14] We have also seen other proposed legislation like the National Defense Authorization Act (NDAA) and bipartisan Deepfakes Task Force Act encourage similar actions. Even internationally, the EU's AI Act and UK Safety Bill also work towards similar goals; this is all positive and helps standardize the practice to mark something in an interoperable and transparent manner if it is created by Gen AI.

Third, the C2PA open standard and digital content provenance is not without challenges. The two largest are education and adoption, but I do believe the government can help with both. On education, Federal and even state/local levels of government can help raise awareness on what content credentials are and are not through larger media literacy campaigns. This will help ensure we have a well informed audience as they begin to proliferate. In addition, many platforms should adopt content credentials but may not have the right aligned incentives to do so. Government can create the right forums, ask the right questions and line up the right incentives in which adoption of transparency standards and best practices are rapidly implemented across industries.

While much of the discussion has rightly been focused on how to protect the public from the negative effects of image deception, we also think there are incredible benefits that the government can realize from authenticated, transparent images and videos. The government should consider how it could use Content Credentials to authenticate its own communications to constituents, helping mitigate risks of synthetic deceptions mimicking government agencies or officials.

Further, as the government deploys funds, aid, and resources across the country and world for development, crisis response, or humanitarian efforts, authentic images with verified provenance

[14] The White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," *The White House*, last modified October 30, 2023, accessed November 4, 2023, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

can help to support government decision making. This is especially true in urgent, non-permissive or non-presence situations, for remote documentation and program monitoring.

Using authenticated images for trusted documentation and program oversight could dramatically cut risk and costs, as well as increase the speed and efficiency of assistance delivery. This is one example of why the government should consider being a first mover, in addition to helping educate and raise awareness to the general public on why transparency and authenticity in digital content are essential. This approach is already being used by hundreds of the largest companies in the world, and, from Truepic's experience, I can confidently say it will dramatically lower costs, save time, and increase speed and efficiency of operations. It will also begin to normalize transparency in digital content, which is an essential protection for decision makers of all levels.

In our estimation, the C2PA open standard and Content Credentials are the most scalable and usable approach for a more transparent and authentic internet. However, we do recognize that there may be other approaches, such as watermarking or detection.We would also encourage the committee to examine all mitigation strategies and urge the government and private sector to begin testing, adopting and implementing approaches to mitigate the risks that accompany undisclosed hyperrealistic AI-generated content and synthetic media.

In closing, I would like to thank this committee for its time and opportunity to present how the C2PA open standard, Truepic, and Content Credentials can help to mitigate some of the risks inherent in the Generative AI world that is rapidly taking shape before our very eyes.

Thank you,

Mounir Ibrahim
Executive Vice President, Public Affairs and Impact

truepic