truepic

U.S. Senate AI Insight Forum:
Transparency, Explainability, Intellectual Property, & Copyright

Written Statement by Mounir Ibrahim, Executive Vice President, Truepic
November 29, 2023

Thank you Leader Schumer, Senators Heinrich, Rounds, and Young, and the other distinguished members of the Senate for the opportunity to submit this statement and participate in today's AI Insight Forum on Transparency, Explainability, Intellectual Property, & Copyright. My name is Mounir Ibrahim, Executive Vice President of Truepic, a technology company focused on transparency and authenticity in digital content since 2015.

Human existence has digitized. 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 as entire industries such as insurance, lending, real estate, peer-to-peer commerce and others rely on images and videos for everything from the security of their operations, to brand integrity. Our reliance on digital information, coupled with a rapid rise in synthetic content, means that transparency in what is real, what is edited, and what is AI-generated has become imperative.

It is our opinion that without scaled transparency in the digital content we all rely on, our economies, democracies, and societies are **at risk**. We are encouraged to see general agreement that transparency in digital content is critical for the internet. Recent domestic and global initiatives ranging from the NDAA and White House's Executive Order, to the EU's AI Act and UK Safety bill, highlight this trend toward a more transparent, authentic internet.

However, I would like to emphasize that to successfully scale transparency in digital content and mitigate the challenges we face online, the standards and technical mechanisms we use must be **interoperable**.

Interoperability

In our estimation, interoperability is one of the most critical pillars for a more transparent and authentic internet. Digital content travels at speed across platforms. Rarely does it remain only on one platform. The velocity and touchpoints are complex; media is created, compressed, edited, uploaded, downloaded, changed and re-posted on different platforms daily, hourly, and by the minute, all over the world. For content transparency to work, each system, platform, browser, and device must be able to read critical information about the original content and add its own relevant information along the way in a chain of custody. This information could include metadata like the creation system, if it is coming from a physical camera (i.e. DSLR), native phone camera, application, or if generated by an AI platform (DALL-E, Stability, Firefly), also the date, time, edits, and other data for the content consumer to review.

On their own, closed systems of transparency or disclosure, like watermarking, will not be sufficient to address the transparency challenge at internet scale as interoperability is at direct odds with their efficacy. Watermarking technology can be incredibly effective and robust in closed systems where a platform - and *only* that platform - has the decoder to read the watermark. However, when that piece of content inevitably moves away from that closed system, its watermark will likely be illegible to other systems. If the decoder is made public, it becomes a significant threat vector as bad actors could add or remove the watermark from content. This re-fuels the liar's dividend,[1] casting doubt on what is real vs. fake, and brings us back to the status quo: the inability to decipher synthetic from authentic, or anything in between.

While not a silver bullet, an **interoperable open standard** that any organization can adopt, contribute to, and leverage is the best option for transparency at internet scale because it is structured to support the dynamic nature of digital content.

C2PA Open Standard

An open technical standard is available and in use thanks to the work of the Coalition for Content Provenance and Authenticity (C2PA). Founded in 2021, the C2PA is a coalition of organizations, including Truepic, Adobe, Microsoft, Sony, Intel, the BBC, Publicis Groupe, Witness, Arm and many others, committed to transparent, authentic provenance in digital content. 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 - image, video, audio and others.

The basic concept of provenance is to use cryptographic hashing and digital signatures to attach the metadata or history of a piece of digital content directly to the file itself. The C2PA standard is structured in a way that makes it **interoperable** - meaning that 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. 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 when they are reviewing its credentials. The transparency information is passed on to content consumers in the form of visual indicators - known as Content Credentials[2] - which can be compared to a nutrition label on food products. Viewers can tap on the overlaid icon to view the Content Credentials on any piece of content.

Point of Creation Deployment

Truepic specializes in what we believe to be the optimal deployment of the interoperable transparency standard: securely **at the point of creation.** Our technology applies C2PA Content Credentials - cryptographically hashing and signing for transparency - at the instant of capture or

[1] 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
[2] ContentCredentials.org: https://contentcredentials.org/, accessed November 3, 2023

generation. This is critical because only at the moment of capture or creation can a system ensure the highest integrity in the data being embedded into the file.

Content Credentials for Synthetic Media

C2PA Content Credentials are being used daily to add transparency to AI-generated outputs. Truepic worked with Revel.ai and Nina Schick to release the world's first transparent "deepfake" video with Content Credentials and with Hugging Face[3] to democratize Content Credentials to any user working with open source models. Adobe launched the same capability in Adobe Firefly and its creative suite.[4] Microsoft's Bing Image Creator also applies Content Credentials to outputs from Open AI's DALL-E.[5] Stability AI has also announced the implementation of Content Credentials in its AI API.[6] Other experts in the field recognize that Content Credentials are a necessity for generative outputs. The Partnership on AI's Responsible Practices Framework for Synthetic Media 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.[7]

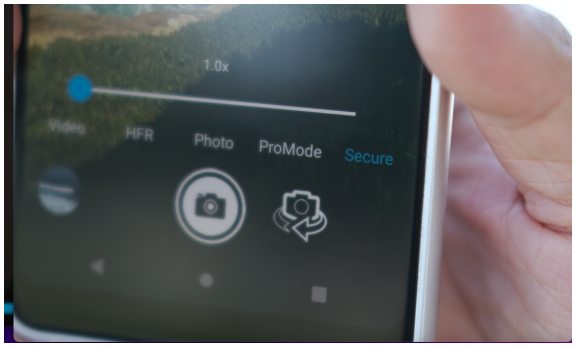
"Mirror of Reflection," First transparent Deepfake with Content Credentials (upper right hand), April 2023.

Content Credentials for Authentic Media

As synthetic media proliferates online, the importance of identifying authentic media will also grow. Content Credentials can also be applied to authentic captures from any smartphone, something Truepic has focused on since its inception. The C2PA standard is built with a privacy-first mindset that allows an opt-in functionality for authentic captures, giving the creator control of what information to include. Truepic **does not** believe Content Credentials for authentic captures should be the default. Rather, we strongly believe there should be an **option** that allows content creators to choose whether or not to turn on Content Credentials in their

[3] 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
[4] "Learn about Content Credentials in Photoshop," accessed November 3, 2023, https://helpx.adobe.com/content/help/en/photoshop/using/content-credentials.html
[5] Microsoft Content Credentials .https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/content-credentials, accessed November 19
[6] "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
[7] "PAI's Responsible Practices for Synthetic Media," *Partnership on AI - Synthetic Media*, accessed November 3, 2023, https://syntheticmedia.partnershiponai.org/

preferred application or directly on their native device (i.e. secure mode vs. portrait or video mode).



Secure mode with authentic capture proven on Qualcomm Chipsets and smartphones[8]

With security and privacy in mind, Content Credentials can be foundational for securing what is authentic. Truepic and Microsoft deployed the first C2PA compliant documentation platform in Ukraine in a program called "Project Providence.[9]" This implementation adds Content Credentials **at the point of capture** as teams work to document destruction in Ukraine. A USAID implementer, the Anti-Corruption Headquarters (ACHQ), successfully used the platform to document the destruction of over 600 cultural heritage sites throughout the conflict over the last year. According to ACHQ, prosecutors in Ukraine have since opened at least 10 different legal investigations using the authenticated images from this secure camera with all provenance details available.

The 'On-Device' Future of Content Credentials for Authentic and Synthetic Media

Currently smartphones capture an estimated 90%[10] of all digital photos globally. Soon our smartphones will also become the most used generative platforms as models are pushed to run locally on devices. Given that roughly 85% of the world's population has access to a smartphone, transparency in both authentic and synthetic media creation directly on handheld devices is critical. Truepic, in partnership with Qualcomm, developed this capability on the Snapdragon 8 Gen 3 chipset in which Content Credentials can be added to either authentic or synthetic media coming from within a Trusted Execution Environment (TEE), the most secure location on the device. This partnership serves as a blueprint by which media created on a smartphone can be sent to any C2PA compliant website, platform, phone, or browser with transparency.

Challenges: C2PA Adoption and Education

We believe that the C2PA open standard is the most suitable, feasible and interoperable approach to scaling transparency in both synthetic and non-synthetic content across the internet. However,

[8] "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

[9] Project Providence: https://www.projectprovidence.io/, accessed Nov 3, 2023

[10] "Almost All Photos Are Now Taken on Smartphones, According to Study," PetaPixel, June 20, 2023, accessed Nov 17, 2023 https://petapixel.com/2023/06/20/almost-all-photos-are-now-taken-on-smartphones-according-to-study/#:~:text=According%20to%20research%20carried%20out,57%2C246%20photos%20taken%20per%20second

it is not without challenges. I would highlight two of the most significant areas: Adoption and Education.

- *Adoption - A Critical Priority:* Similar to any open standard, the C2PA needs to be widely adopted to be most effective. Without wide adoption, inadvertent edits may disrupt the provenance trail. While we are seeing significant first movers leveraging the standard - wider adoption must be encouraged across the ecosystem. Furthermore, the platforms on which the world most regularly views digital content - social media platforms - must agree to ingest C2PA-tagged content and relay those Content Credentials to viewers. Currently, most social media platforms strip data out of digital content posted on their sites. However, when media files have C2PA Content Credentials it is for the explicit, net-positive purpose of transparency. If an authentic capture has Content Credentials, it means *the creator knowingly opted* to add that transparency to their work. If synthetic content has Content Credentials, it means that the generative platform *adhered to best practices* and committed to transparently mark its outputs. In either case, social media platforms, browsers, and other informational platforms **must not strip C2PA** and block these net-positive efforts for scaled transparency.

- *Education & Explainability - A Cross-Sector Effort:* Even with wide-scale C2PA adoption, another, perhaps larger, challenge still exists. As Content Credentials proliferate, we must educate and explain their meaning to the general public. In our estimation, explaining what they mean is as important as explaining what they *do not* mean. Content Credentials *are not* instant indicators of truth or falsehood. Rather, they are a tool to help us evaluate what we see and hear online. They are a prompt to apply our media literacy skills and look at the provenance of digital content before making a decision of consequence, similar to how a shopper may look at a nutrition label before purchasing an item at a local supermarket. Government can play a significant role here.

Government can be a first-mover and adopter of C2PA to make Content Credentials more visible to the public and highlight how they can add transparency to official communications and media. Government leadership can also help to establish a normalcy around transparency signals in digital content. Additionally, government agencies such as the NSF, Dept of Commerce, Dept of Education, Dept of State and many others can fund research on the optimal methods to display and explain Content Credentials to the general public. Initial studies from researchers at Oxford,[11] and University of Washington,[12] are great starting points to build a broader knowledge base for industry, government, and media to optimally deploy Content Credentials across the internet. This research can also be built into curricula to enhance understanding across society.

[11] Cassidy Bereskin, "Understanding the Efficacy of Provenance Interventions for Tackling Misinformation," last modified March 26, 2023, accessed November 20, 2023, https://osf.io/4a2fw/.
[12] K. J. Kevin Feng et al., "Examining the Impact of Provenance-Enabled Media on Trust and Accuracy Perceptions" (arXiv, September 10, 2023), accessed November 20, 2023, http://arxiv.org/abs/2303.12118.