truepic

Senate Committee on Commerce, Science, and Transportation's Subcommittee on Consumer Protection, Product Safety and Data Security

Protecting Consumers from Artificial Intelligence Enabled Fraud and Scams

Written Testimony Submitted by

Mounir Ibrahim
Chief Communications Officer and Head of Public Affairs
Truepic

November 19, 2024

Thank you, Chairman Hickenlooper, Ranking Member Blackburn, and esteemed members of this subcommittee. My name is Mounir Ibrahim, and I represent Truepic—a technology company dedicated to providing essential transparency and authenticity tools for the digital content we encounter daily.

The importance of understanding the origins and authenticity of digital content was deeply impressed upon me during my time as a Foreign Service Officer with the U.S. Department of State, where I worked on various global conflicts, including serving on the ground at our Embassy in Damascus, Syria. My experiences as a diplomat exposed me to the urgency of verifying the authenticity of digital content at the highest levels of national and global security decision-making.

I fundamentally believe deciphering what is human-created from AI will become one of the most pressing challenges across all aspects of life.

At Truepic, we recognize that our increasingly digitized lives rely on digital content for decisions—personal, business, and governmental. From insurance claims and banking audits to social media feeds, online profiles, and even images from conflict zones - digital content shapes the decisions we make every day.

I applaud the subcommittee's focus on the impact of digital content on people and consumers in today's AI age. I also thank you for your continued leadership in supporting transparency online.

Threat Landscape

You are undoubtedly aware of the rapid growth of AI platforms, many of which are publicly available, open-sourced, and easy to access. While much attention has rightly been given to how synthetic media and deepfakes could impact elections, mischaracterize leaders, and manipulate markets, I would like to briefly address the threat to local communities and everyday individuals who lack the resources to quickly and easily debunk false content.

Local schools, community leaders, businesses, and law enforcement face significant challenges in distinguishing human-created from synthetic content online, leaving them vulnerable as bad actors exploit easily accessible AI tools with little or no technical expertise.

This dynamic is most clearly seen in the prevalence and alarming rise of non-consensual pornography, often targeting young women, even minors. A recent study across 10 countries performed by researchers in Australia and the United States found that 2.2% of respondents were victims of deepfake pornography, while 1.8% admitted to creating or sharing it. Meanwhile, Catphishing and Sextortion scams are rapidly increasing and more often powered by AI technology.

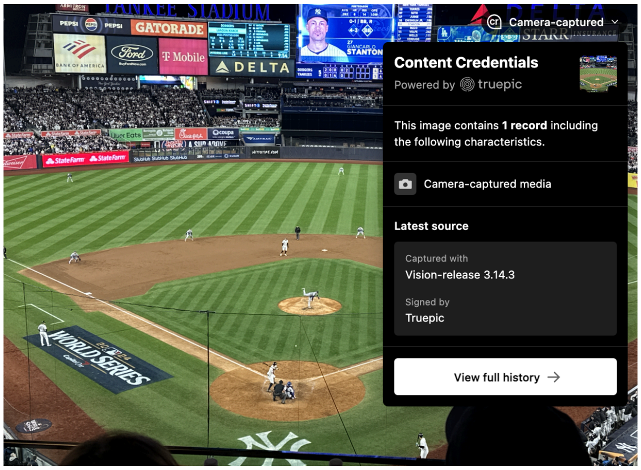
I fear we are witnessing the early stages of AI being weaponized against local communities and individuals who lack the resources to defend themselves. AI-driven visual deception is rapidly expanding beyond non-consensual pornography into local politics, schools, and business fraud. In New York and Maryland, for example, deepfake videos falsely depicted school leaders making offensive and racist remarks, uprooting communities and making national news. In Louisiana, bad actors on platforms like 4Chan have shared tools to create deepfakes targeting judges, prosecutors, and defendants during locally streamed parole board hearings.

We know that businesses also face the same challenges. It has been reported that deepfake phishing scams grew by 3000% and deepfake incidents in the fintech sector increased by 700% last year alone. These trends harm consumers, small business owners, and jeopardize America's economic competitiveness.

What Can Be Done?

The reality is that there is no single solution to halt the growing trend of visual deception. However, there are effective strategies to mitigate the threat, elevate human-created content, and enhance transparency across the internet. I will defer to my esteemed colleague on the panel to discuss the capabilities of the detection of AI content, but one key approach I want to emphasize to this committee is the importance and potential of digital content provenance.

Content provenance securely and cryptographically attaches information (metadata), known as Content Credentials, to photos, videos, or audio so consumers can understand where it came from, how it was created, and whether it has been edited. For example, Content Credentials can show if something was made by a camera or generated by AI, when and where it was created, and any changes it has gone through. This helps people understand what they see online by providing a clear record of its origins and history.



This image contains 1 record including the following characteristics.

Captured with
Vision-release 3.14.3

Signed by
Truepic

Image of 2024 World Series with Content Credentials

This is not hyperbole. Today, Content Credentials are being used by many companies and leveraged on some of the world's most-used social media sites (LinkedIn, YouTube, Instagram, Facebook, etc.). The approach is driven by an open standard developed by the Coalition for Content Provenance and Authenticity (C2PA), of which Truepic is a proud steering committee member, along with Microsoft, Adobe, Google, OpenAI, BBC, and many others.

<u>How Is It Being Used?</u>

I would like to share how Truepic is leveraging Content Credentials and how that is helpful to consumers. First, let me begin with how the private industry has deployed Content Credentials as a necessary pre-requisite to operating in the AI world.

- Business credentialing and re-credentialing are quickly becoming one of the leading industries to embrace content provenance for verifying buyers of credit reports. Without verifiable and transparent content, companies face significant risks and become more vulnerable to AI-driven fraud. Provenance technology enables our partners like Equifax, Transunion, and Dun & Bradstreet to securely transform the verification process.
    - Content Credentials protect consumers by enabling our partners to digitally verify the authenticity of organizations purchasing credit reports. This helps ensure that potentially sensitive credit information is shared only with authorized and legitimate businesses, helping safeguard consumers from AI fraud and deception.

- Insurance is another critical industry that has digitized its processes and is subject to significant fraud that raises rates and complicates processes for consumers. With partners such as EXL Service, USAA, Jewelers Mutual, and many more, our partners are leveraging content provenance to streamline claims processing, underwriting, reduce fraud, and improve customer experiences. Other partners, like Palomar Specialty Insurance, deploy content provenance to address natural disasters, speeding up the process for and supporting victims through authenticated content.

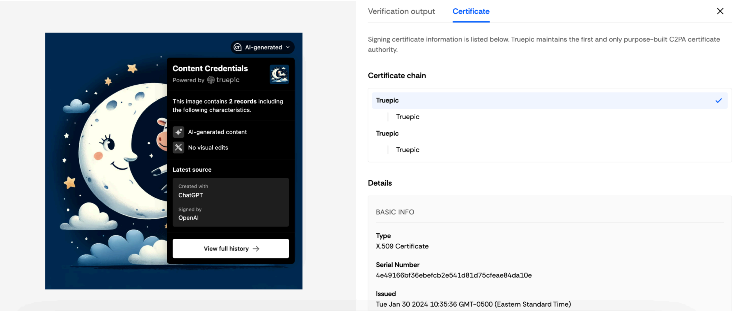
- ○ Content Credentials protect consumers in insurance by ensuring claims and underwriting are supported by genuine and verifiable evidence, preventing fraud that can inflate premiums for everyone. This benefits policyholders by allowing them to receive fair and timely resolutions.

We are excited to extend this impact to other consumer-focused areas like product recalls through partnerships with organizations like Sedgwick. Verifiable information through Content Credentials ensures that owners and victims of defective products can quickly prove possession and malfunction, allowing them to be compensated more efficiently and without unnecessary red tape, enhancing consumer protection.

Public Facing Benefit

Let me address how Content Credentials are being deployed to help power a more authentic and transparent online experience essential to help stem the deception of people online.

OpenAI's ChatGPT, including DALL·E, is one of the most prominent AI text-to-image generation platforms. It has implemented stronger safeguards to combat misuse and prevent deception, an effort we are proud to support. Truepic plays a key role in this initiative by providing the Certificate Authority that backs the C2PA signature on AI-generated images, ensuring transparency and accountability. Content Credentials on OpenAI outputs help consumers because the Truepic certificate will be recognized by major social media platforms like LinkedIn, Instagram, or Facebook, and the consumer will be alerted the image is AI-created.



This image contains **2 records** including the following characteristics.

AI-generated content

No visual edits

Latest source

Created with
ChatGPT

Signed by
OpenAI

Signing certificate information is listed below. Truepic maintains the first and only purpose-built C2PA certificate authority.

Certificate chain

Truepic ✓
Truepic
Truepic
Truepic

Details

BASIC INFO

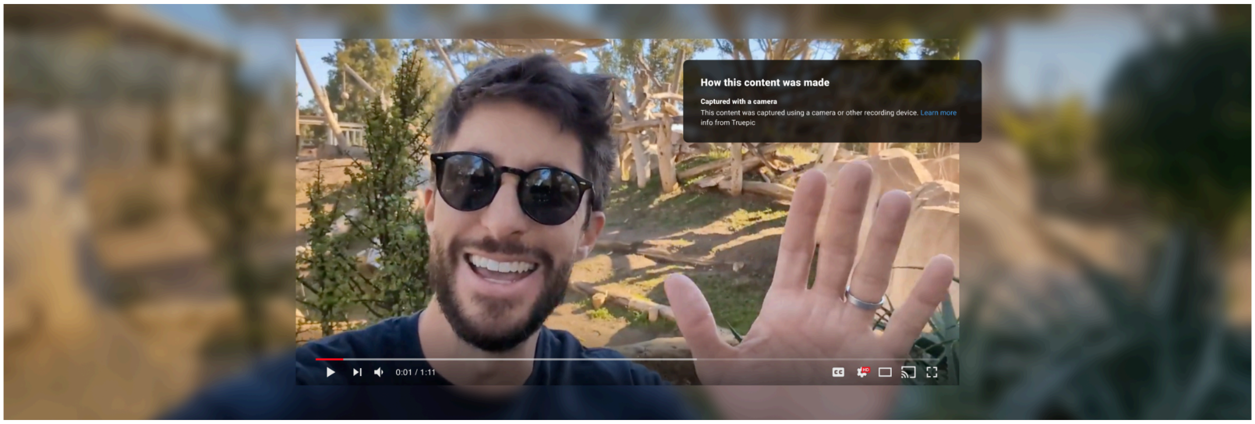
Type
X.509 Certificate

Serial Number
4e49166bf36ebefcb2e541d81d75cfeae84da10e

Issued
Tue Jan 30 2024 10:35:36 GMT-0500 (Eastern Standard Time)

Image created on ChatGPT with C2PA Content Credentials backed by Truepic's Certificate Authority

We are also working to help power any organization that wants to capture and share authentic video on YouTube. Last month, we produced the first human-created video with Content Credentials on YouTube, which much like the prior example, allows consumers to understand the origins of the video they are watching. This capability can be extended to any application that captures videos and places them on YouTube.



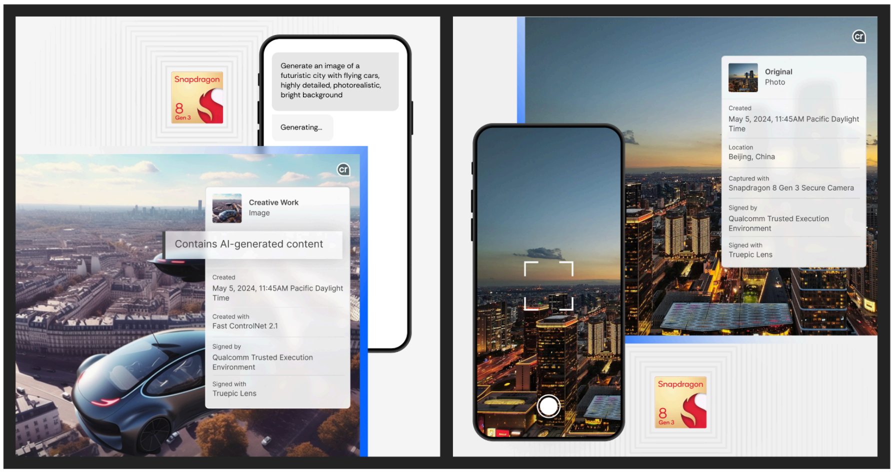
How this content was made

Captured with a camera

This content was captured using a camera or other recording device. Learn more info from Truepic

YouTube displays Content Credentials on authentically-produced videos

Another exciting example of this technology creating transparency for consumers lies in our smartphones. With AI increasingly accessible on mobile devices, we're proud to partner with Qualcomm to embed Truepic technology directly into Snapdragon chipsets. This enables on-device AI content to be tagged with Content Credentials. At the same time, creators can also choose to tag authentic images and videos, ensuring transparency in digital content while safeguarding privacy.



Generate an image of a futuristic city with flying cars, highly detailed, photorealistic, bright background

Creative Work
Image

Contains AI-generated content

Created
May 5, 2024, 11:45AM Pacific Daylight Time

Created with
Fast ControlNet 2.1

Signed by
Qualcomm Trusted Execution Environment

Signed with
Truepic Lens

Original
Photo

Created
May 5, 2024, 11:45AM Pacific Daylight Time

Location
Beijing, China

Captured with
Snapdragon 8 Gen 3 Secure Camera

Signed by
Qualcomm Trusted Execution Environment

Signed with
Truepic Lens

Qualcomm and Truepic developed the first chipset to power digital content transparency across smartphones worldwide

Challenges

While we are highly encouraged by this progress, it is important to acknowledge that significant challenges persist.

- Adoption: For the open specification to reach maximum efficacy, we need the various components of the internet to adopt and implement, this includes browsers, platforms, CDNs (Content Delivery Networks), and OEMs (Original Equipment Manufacturers). Without widespread adoption, Content Credentials can be easily lost when content moves to non-compliant platforms. The C2PA expects the International Standards Organization (ISO) to approve the specification as an open standard in the coming months, and we anticipate there will be further adoption; however, much more needs to be done.

- Education: We also need to work together - industry, government, and civil society to educate all stakeholders on what Content Credentials mean and, perhaps more importantly, do not mean. They are not meant and should not be considered stamps of blind trust. Rather, they are indicators that more information on that piece of content is available so that consumers can make more accurate decisions based on digital content. We believe this is essential and the C2PA, supported by a generous grant from Microsoft and OpenAI, has been working hard to accelerate educational efforts on the specification and approach.

Moving Forward & Recommendations

We are dedicated to working with partners and the C2PA community to ensure Content Credentials empower more informed decisions online. We also believe that government can be critical in advancing a more transparent internet, especially with the following:

- Education & Funding: We believe this is the most immediate area in which the government can support various education initiatives and research institutions examining how provenance and Content Credentials can be most effective in supporting consumers.
- Engagement: Hearings like this with policymakers, industry leaders, and civil society are essential to raising awareness. I would encourage looking at provenance and transparency beyond just a safety measure, and more importantly, as an ***opportunity*** that unlocks significant government and private sector efficiencies.
- Existing Recommendations & Research: In line with the bipartisan AI Insight Roadmap, we encourage this committee to consider how establishing provenance of digital content, for both synthetic and non-synthetic content, can be beneficial. NIST is also performing a critical role as it looks at how transparency in AI can be more readily available.

Transparency is essential as digitization accelerates and the line between human-created and synthetic content blurs. Without it, consumers face greater deception, businesses endure rising fraud, and governments risk miscalculation and exploitation. These challenges are significant, but with collaboration and the awareness driven by hearings like this, I am confident we can build a more transparent and authentic digital future.

Thank you for considering this testimony.