MY WEIRD PROMPTS

Podcast Transcript

EPISODE #372

Proving Reality: Fighting the Liars Dividend with C2PA

Published January 30, 2026 • Runtime: 26:29

https://myweirdprompts.com/episode/ai-deepfakes-truth-verification/

EPISODE SYNOPSIS

As generative AI makes it easier than ever to fabricate reality, we are entering the era of the "liars dividend"—a world where any piece of real evidence can be dismissed as a computer simulation. In this episode, Herman and Corn dive deep into the technical and legal frameworks struggling to preserve the truth, from the Content Authenticity Initiative (CAI) to the hardware-level security chips in professional cameras. They explore how cryptographic "nutrition labels" for images work, whether your smartphone can actually be trusted in court, and the growing danger of a "technology gap" that could create a two-tiered system of truth. This is a must-listen for anyone concerned about the future of evidence, journalism, and our shared sense of reality in 2026 and beyond.

DANIEL'S PROMPT

Daniel

I'd like to discuss content authenticity in the context of the rapid rise of generative AI. Many people may soon find themselves facing allegations that their real content is actually AI-generated. While we've discussed invisible fingerprinting, the lack of a fingerprint doesn't necessarily prove something isn't AI-generated. This is where content authenticity standards come in. The Content Authenticity Initiative (CAI) has formed an industry consortium to develop these standards, and apps like ProofMode allow users to embed metadata and verification at the software level. There is also hardware-level verification with C2PA-approved devices. My question is: do you think software-level verification is only a stepping stone toward hardware-level verification? How much legitimacy can be attached to software-level authenticity verification? Is hardware-level signing the gold standard for tamper-proof content? In high-stakes environments like law enforcement, is hardware-level authenticity certification necessary or already being implemented for body cams?

TRANSCRIPT

Corn

Have you ever had that unsettling feeling that the ground is shifting under your feet? Not literally, though living here in Jerusalem, we do get the occasional tremor. I am talking about the conceptual ground. The feeling that the very idea of evidence is evaporating. I was thinking about this all morning after listening to the audio our housemate Daniel sent over. He had this nightmare scenario with a leaky roof and a landlord who tried to claim a photo of mold was A-I generated. It sounds like a plot from a sci-fi movie, but it is actually just Tuesday in twenty twenty-six.

Herman

Herman Poppleberry here, and Corn, you are hitting on the exact nerve that has been twitching in the tech community for the last couple of years. Daniel is really onto something with this prompt. It is what researchers call the liars dividend. It is the idea that as soon as deepfakes and generative A-I become common, anyone caught doing something wrong can simply point to the evidence and say, that is not real, it is just a computer simulation. It creates a vacuum of truth where the default assumption shifts from seeing is believing to seeing is suspicious.

Corn

It is a terrifying shift. If you cannot prove a photo of a leaky roof is real, how do you prove anything? Daniel mentioned the Content Authenticity Initiative, or C-A-I, and this standard called C-two-P-A. I have seen that little C-R icon popping up on more images lately. It is starting to feel like the nutrition label for the internet. But I want to really dig into the mechanics today. Is this actually going to save our sense of reality, or is it just another layer of digital cat and mouse?

Herman

Well, the C-A-I is a massive consortium. You have got Adobe, Microsoft, Truepic, and even some of the big camera manufacturers like Sony and Leica. The goal is to create a permanent, tamper-evident record of where a piece of content came from and how it was changed. It is called provenance. Think of it like a digital chain of custody that is baked into the file itself.

Corn

Okay, so let us start with what Daniel was using. He mentioned ProofMode. That is a software-level tool. If I am using an app like that on my phone to take a photo of a leaky roof, what is actually happening under the hood? How is that different from just a regular J-P-E-G with a timestamp?

Herman

That is a great place to start. A regular photo has what we call E-X-I-F data. That tells you the shutter speed, the date, and maybe the G-P-S coordinates. But the problem is that E-X-I-F data is incredibly easy to fake. I could open a photo in a basic hex editor and change the date to nineteen ninety-nine if I wanted to. Software-level verification like ProofMode goes much deeper. When you take a photo in an app like that, it captures a massive burst of sensor data and environmental metadata. It looks at the cell towers you are connected to, the Wi-Fi networks in range, the barometric pressure, and even the light sensor readings.

Corn

So it is creating a fingerprint of the physical moment, not just the image.

Herman

Exactly. And then it takes all that data and creates a cryptographic hash. A hash is basically a digital signature that is unique to that specific set of data. If even one pixel in that photo is changed, or if one digit of the G-P-S coordinate is altered, the hash will no longer match. The app then signs that hash with a private key that is unique to your device. This is where we get into the Federal Rules of Evidence that Daniel was curious about. In the United States, for example, Federal Rule of Evidence nine hundred two covers self-authenticating evidence.

Corn

Right, I remember reading about that. Usually, you need a witness to stand up in court and say, I took this photo and it is accurate. But nine hundred two allows certain records to be admitted without a witness if they have a certified digital signature.

Herman

Precisely. Rules nine hundred two thirteen and fourteen were specifically added to handle things like forensic copies of hard drives and electronic data. Software like ProofMode is designed to meet that standard. It creates a bundle of evidence that a lawyer can take to a judge and say, this file has not been altered since the moment it was captured, and here is the mathematical proof.

Corn

But here is my concern, Herman. If it is just software, is it not still vulnerable? If I am a sophisticated actor, can I not just spoof the software? Can I not feed the ProofMode app a fake video stream or fake G-P-S coordinates from a rooted phone?

Herman

You have hit the nail on the head. That is the fundamental limitation of software-only verification. You are trusting the operating system. If the Android or i-O-S kernel is compromised, or if you are running the app in an emulator, you can feed it whatever lies you want. The software will faithfully sign those lies, and you end up with a perfectly verified, high-integrity fake. This is why many experts see software-level tools as a vital first step, but not the final destination.

Corn

Which leads us to the hardware-level verification. Daniel asked if hardware is the gold standard. We are starting to see this in professional cameras now, right? I know Leica was the first to really jump on this with the M-eleven-P.

Herman

Yes, the Leica M-eleven-P was the pioneer. It has a dedicated security chip inside the camera body. When the light hits the sensor and the image processor creates the raw file, that security chip immediately signs the data. This happens before the data even reaches the memory card or any external software. It is what we call a hardware root of trust. Sony followed suit not long after. They released firmware updates for the Alpha one, the Alpha seven-S mark three, and the Alpha nine mark three to support C-two-P-A.

Corn

So in those cases, the camera itself is saying, I am a physical device, I saw this light at this time, and I am signing it with a key that is physically burned into my silicon. You cannot spoof that by rooting a phone because the key never leaves the secure enclave of the camera.

Herman

Exactly. It creates a much higher barrier for fraud. To fake a hardware-signed image, you would essentially have to physically hack the chip or build a sophisticated optical rig to project an A-I generated image onto the camera sensor in real-time. It is not impossible, but it is orders of magnitude harder than clicking a button in a software suite.

Corn

That makes sense for a professional photographer or a journalist in a war zone. But what about the average person? Most of us are not carrying five-thousand-dollar Leicas. We are using our phones. Is this coming to the hardware in our pockets?

Herman

It is already starting. Qualcomm and Apple have been members of these consortia for a while. Modern smartphones already have secure enclaves and T-E-Es, which are Trusted Execution Environments. These are the same chips that handle your Face I-D and your credit card information for mobile payments. The infrastructure is there. The next step, which we are starting to see in twenty-twenty-six, is integrating the camera pipeline directly into that secure enclave.

Corn

So eventually, every photo I take on my phone could have a C-two-P-A seal of approval by default. But that brings up a huge privacy question, does it not? If every photo is cryptographically linked to my specific device and my location, am I not just creating a perfect tracking tool for myself?

Herman

That is the big tension. The C-A-I has tried to address this by making the metadata optional or redactable. You can prove a photo is real without necessarily revealing exactly where you were or who you are. You can use what is called a zero-knowledge proof or just a partial disclosure. For example, you could prove the photo was taken in Jerusalem on January thirtieth without revealing the exact street address. But you are right to be cautious. Any system built to prove truth can also be used to enforce surveillance.

Corn

It is a delicate balance. I want to go back to the high-stakes environments Daniel mentioned. Law enforcement and body cams. This seems like the most obvious use case. If a police officer is involved in a shooting, the public needs to know that the footage has not been edited or deepfaked by the department. Is this actually happening yet?

Herman

It is in the transition phase. Companies like Axon, which is the big player in the body cam space, have joined the C-two-P-A steering committee. They are working on integrating these standards into their entire ecosystem. Think about the chain of custody. Right now, we rely on a secure upload from the camera to a cloud server like Evidence-dot-com. But with C-two-P-A, the footage is signed the second it is recorded. Even if it sits on a thumb drive in a drawer for a week, you can verify its integrity later.

Corn

That seems like it should be mandatory. If we have the technology to prove a video is untampered, why would we ever accept anything less in a courtroom?

Herman

Well, there is the cost and the legacy hardware problem. Police departments are not known for being quick to upgrade their tech. There are hundreds of thousands of older body cams in service that do not have the hardware to support this. We are likely looking at a decade-long transition. But in twenty-twenty-six, we are seeing the first precincts starting to require C-two-P-A compliance for all new equipment. It is becoming a benchmark for transparency.

Corn

I wonder about the defense side of that, though. If hardware-level signing becomes the gold standard, does that mean any evidence that does not have it is automatically seen as fake? If I am a bystander who records police misconduct on an older phone that does not support C-two-P-A, will a jury just ignore it because it lacks the digital seal of approval?

Herman

That is a brilliant point, Corn, and it is something that civil liberties groups are very worried about. It is called the technology gap. If we create a world where only verified content is trusted, we might inadvertently silence the people who cannot afford the latest verified devices. It could create a two-tiered system of truth. The official story is verified by expensive hardware, while the grassroots story is dismissed as unverified noise.

Corn

We saw something similar with the transition to high-definition video. People started to think that if a video was grainy, it was somehow less authentic, even though graininess is often a sign of a real, low-light situation. Now we are moving from visual aesthetics to cryptographic aesthetics.

Herman

Exactly. And this is why software-level tools like ProofMode are so important. They provide a middle ground. They might not be as bulletproof as a dedicated security chip in a Leica, but they are accessible to anyone with a smartphone. They allow a regular person to add a layer of credibility to their evidence without needing a specialized device. It is about democratizing the ability to prove reality.

Corn

I think we should talk about the role of the platforms here. Adobe has been the big driver behind this, which is interesting because Photoshop is the tool everyone blames for fake images in the first place. It is like the company that made the lock is now making the key.

Herman

It is a smart move by them. They have integrated Content Credentials directly into Photoshop. When you edit an image, the software keeps a log of what you did. If you used an A-I tool to fill in a background or change someone's hair color, that is recorded in the metadata. When you export the file, you can attach those credentials. Then, when someone sees that image on a news site or social media, they can click that little C-R icon and see exactly how the image was altered.

Corn

But that only works if the platforms support it. If I upload a verified photo to a social media site and that site strips out all the metadata to save space, the chain of trust is broken.

Herman

That has been the biggest hurdle. Most social media platforms have historically stripped E-X-I-F data for privacy and storage reasons. But we are starting to see a shift. Some platforms are beginning to recognize C-two-P-A manifests and display that authenticity badge. They realize that in a world of A-I hallucinations, being the platform that can prove what is real is a huge competitive advantage.

Corn

It is basically a new form of the blue checkmark. Instead of verifying the person, we are verifying the pixels. But we saw how the blue checkmark system fell apart when it became a pay-to-play scheme. Is there a risk of that here?

Herman

The difference is that C-two-P-A is an open standard. It is not owned by one company. Anyone can build a tool to verify these manifests. You do not have to pay a monthly fee to a billionaire to have your photos cryptographically signed. You just need a device or an app that follows the protocol. That open-source nature is what gives me hope that it will not just become another corporate gatekeeping tool.

Corn

I want to go back to Daniel's specific situation. Let us say he uses ProofMode to document the mold from the leaky roof. He goes to court against his landlord. The landlord says, that is A-I. Daniel's lawyer pulls up the ProofMode bundle. How does a judge, who might not be tech-savvy, actually interpret that? Do we need a new breed of digital forensic experts to explain these hashes to a jury?

Herman

We already have them, but their jobs are about to get a lot busier. In a case like that, the lawyer would likely use a verification tool. There are websites where you can drop a file, and it will check the C-two-P-A manifest against the public keys of the manufacturer or the software provider. It gives you a green checkmark and a simple explanation. The goal of the C-A-I is to make it so simple that a judge or a juror can understand it without a degree in computer science. It is like a notary's seal. You do not need to understand the chemistry of the ink or the physics of the stamp to know that a notarized document is official.

Corn

That is a good analogy. But a notary's seal can still be forged. It just takes a lot of effort. It feels like we are in this perpetual race where the fakes get better, so the verification has to get stronger, which then challenges the fakers to get even better. Does this race ever end?

Herman

Probably not. It is the nature of security. But what we are doing with C-two-P-A is raising the floor. Right now, the floor is so low that a teenager with a free A-I tool can create a convincing fake. By implementing these standards, we are making it so that creating a convincing, verified fake requires state-level resources. It does not eliminate the problem, but it makes it much harder to do at scale.

Corn

And scaling is the real danger. One deepfake is a problem, but a million deepfakes is a catastrophe for democracy. If we can at least verify the work of professional journalists and official records, we can maintain a baseline of shared reality.

Herman

Exactly. It is about protecting the signal in the noise. One of the most interesting developments I have seen lately is how this is being used in citizen journalism. There is a project called Starling Lab, which is a collaboration between Stanford and U-S-C. They have been using these tools to document potential war crimes. They use a combination of hardware-level signing, decentralized storage on things like Filecoin, and even recording the hashes onto a blockchain.

Corn

That is intense. Using a blockchain to preserve the history of a war zone. It makes sense, though. If you want to create a record that is truly permanent and tamper-proof, you have to move beyond a single server or a single company.

Herman

Right. They call it a framework of trust. You have the capture layer, which is the camera signing the data. You have the storage layer, which is decentralized so no government can delete it. And you have the verification layer, which is the open standard that anyone can check. It is a complete reimagining of how we preserve history in the digital age.

Corn

It makes me think about episode one hundred fifty-one, where we talked about why high-speed internet can still feel slow. It was all about the bottlenecks in the system. In the world of content authenticity, the bottleneck is not the technology, it is the trust. We have the math to prove things are real, but do we have the social structures to believe the math?

Herman

That is the million-dollar question. You can show someone a green checkmark and a cryptographic proof, but if they are determined to believe a lie, no amount of math will change their mind. We are building the technical tools for truth, but the psychological tools are still lagging behind.

Corn

So, to answer Daniel's question, software-level verification is a vital tool for the masses, but hardware-level signing is the necessary backbone for high-stakes truth. We are likely moving toward a world where your phone does this automatically, but we have to be vigilant about the privacy and accessibility implications.

Herman

Well said. It is not a silver bullet, but it is the best shield we have against the coming storm of synthetic content. We are moving from an era of blind trust to an era of verified provenance. It is going to be a bumpy ride, but at least we have a map.

Corn

I think that is a good place to wrap up our core discussion. But before we go, I want to make sure we give some practical advice for people like Daniel who might be facing these accusations right now. If you are in a situation where you need to prove your content is real, what should you do?

Herman

First, if you know you are in a high-stakes situation, like a legal dispute or a sensitive reporting job, start using a verification app immediately. ProofMode is excellent and open-source. There is also an app called Truepic Lens that is very highly regarded in the professional space. These apps create a bundle of metadata that is much harder to dismiss in court than a standard photo.

Corn

And if you are buying new equipment, especially for professional use, check for C-two-P-A support. It is becoming a standard feature in high-end cameras, and it likely won't be long before it is a selling point for smartphones too. Also, keep your original files. Never just keep the version you uploaded to social media. The original file has the most metadata and the highest integrity.

Herman

Great point. And for the listeners who are just browsing the web, start looking for that C-R icon. Adobe has a website called Verify where you can upload any image to see if it has Content Credentials. It is a great way to train your own eye to look for provenance.

Corn

This has been such a fascinating deep dive. Daniel, thanks again for sending this in. It is one of those topics that feels a bit technical until it suddenly becomes very personal, like with your leaky roof. I hope the landlord eventually backed down when he saw you knew your stuff.

Herman

Yeah, nothing scares a difficult landlord like a tenant who understands cryptographic hashes. Before we sign off, I want to remind everyone that we love hearing from you. If you have a weird prompt of your own, or if you have thoughts on the reality crisis we discussed today, head over to myweirdprompts-dot-com and use the contact form.

Corn

And if you have been enjoying the show, we would really appreciate it if you could leave us a review on Spotify or your favorite podcast app. It genuinely helps other curious minds find us in the sea of content. We are on episode three hundred sixty-five now, and it is the support of this community that keeps us going.

Herman

It really does. We have come a long way since those early episodes, and we have got a lot more rabbit holes to explore. Remember, you can find all our past episodes and more information at our website.

Corn

Thanks for joining us in the house today. This has been My Weird Prompts. I am Corn.

Herman

And I am Herman Poppleberry. We will see you in the next one.

Corn

Stay curious, and keep questioning the pixels. Bye for now.

Herman

Bye everyone.