pISSN:2349-5707 Volume:13,Issue 4, April -2026 eISSN:2349-5715

THE LIAR'S DIVIDEND IN THE COURTROOM: WHY FEDERAL RULE OF EVIDENCE 901 MUST ABANDON THE 'REASONABLE JUROR' STANDARD FOR DIGITAL MEDIA

Author: Kurbanaliyev Sardor Alidjanovich
Institution: Tashkent International University
Field of Study: Jurisprudence (Law), 2nd-year student
Phone: +99890-451-53-05 Email: sardorkurbon@gmail.com

Abstract. This article critiques the current evidentiary standards for digital media in United States criminal trials, arguing that the rise of high-fidelity "Deepfakes" has rendered Federal Rule of Evidence 901(a) obsolete. By analyzing the "Liar's Dividend"—the phenomenon where guilty defendants successfully claim real evidence is fake—it demonstrates that the current "low bar" for authentication paradoxically aids the defense while endangering the prosecution. The article reviews the Advisory Committee on Evidence Rules' 2025 decision to delay adopting a specific "Deepfake Rule" and argues this inaction was a fundamental error. It proposes a new "Verified Capture Standard" (VCS) that leverages the 2025 C2PA cryptographic protocols to establish a presumption of authenticity, effectively modernizing the "Chain of Custody" for the algorithmic age.

Introduction. In the 20th century, the "Silent Witness" theory of evidence posited that a photograph or video spoke for itself. It was an objective observer, immune to the lapses of human memory or the bias of a narrator. The camera was viewed as a mechanism of truth—a mechanical eye that captured reality with indisputable fidelity.

In 2026, that witness has become a perjury machine.

The proliferation of Generative Adversarial Networks (GANs) and Diffusion Models has democratized the creation of synthetic reality. Today, a teenager with a laptop can generate audio of a CEO confessing to fraud, or video of a rival committing assault, with a fidelity that defies human detection. While the legal community has long feared the admission of fake evidence (the "False Positive" problem), a more insidious threat has emerged: the **"Liar's Dividend."**[1]

The Liar's Dividend refers to the skepticism that deepfakes inject into the ecosystem of truth. As the public becomes aware that anything can be faked, they begin to suspect that everything is fake. This allows guilty defendants to challenge authentic audio-visual evidence by simply raising the specter of AI manipulation, increasing the burden of proof for the state to impossible heights. When a video of a crime is presented, the defense counsel need only ask, "Can you prove this isn't a deepfake?" If the prosecution cannot prove a negative, reasonable doubt festers.

This article argues that the Federal Rules of Evidence (FRE), specifically Rule 901, are structurally incapable of addressing this crisis. By clinging to the "reasonable juror" standard—which assumes laypeople can distinguish truth from fiction—the courts are inviting a collapse of confidence in the justice system. The "wait-and-see" approach adopted by the Judicial Conference in 2025 was a miscalculation; the technology is moving faster than the common law can evolve.

To understand the legal failure, one must first accept the forensic reality of 2026: passive detection is dead.

For years, forensic experts relied on "artifacts" to spot fakes. Early deepfakes (circa 2019-2022) had tell-tale signs: mismatched lighting, irregular blinking patterns, or spectral

pISSN:2349-5707 Volume:13,Issue 4, April -2026 eISSN:2349-5715

inconsistencies in audio frequencies. Prosecutors relied on expert witnesses to point out these flaws. However, this reliance ignored the fundamental architecture of Generative AI.

The Generator vs. The Discriminator Deepfakes are created using Generative Adversarial Networks (GANs). A GAN consists of two neural networks locked in a zero-sum game:

1. The Generator: Creates a fake image or audio clip.

2. The Discriminator: Attempts to spot the fake. The two networks train against each other billions of times. If the Discriminator spots a flaw (e.g., "the shadows don't match the light source"), the Generator learns from that failure and corrects it in the next iteration. This means that AI is specifically designed to defeat detection. By late 2024, "self-correcting" adversarial models rendered traditional forensic analysis obsolete. If a forensic tool exists to spot a fake, the GAN can be trained against that specific tool until the artifact disappears.[2]

The Audio Crisis: "VoiceClone-X" While video deepfakes grab headlines, the most immediate threat to criminal justice is **audio synthesis**. Unlike video, which requires massive data to render light physics and texture, audio is low-bandwidth. Models like "VoiceClone-X" (a pseudonym for current enterprise tools available in 2026) need only three seconds of reference audio—easily obtained from a voicemail or social media post—to synthesize a perfect clone. These models capture the speaker's unique cadence, breathing patterns, and emotional intonation (prosody). In United States v. Anthony (2025), prosecutors attempted to introduce a voicemail where the defendant threatened a witness. The defense produced a "battle of the experts," with one forensic analyst claiming the audio lacked the "spectral jitter" of a deepfake, and another claiming it was a "high-quality synthesis." The judge, lacking a clear standard, excluded the audio under FRE 403 (confusion of the issues). The result? A likely authentic threat was silenced because the possibility of a fake was too high to litigate.[3]

Federal Rule of Evidence 901 governs the authentication of evidence. Its threshold is famously low. Subsection (a) states:

"To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is."

Historically, this meant a witness simply had to say, "Yes, that's what the scene looked like." The jury would then decide how much weight to give it. This structure rests on a fallacy: that the **"naked eye"** is a reliable detector of forgery.

The leading pre-AI precedent, United States v. Vayner (2nd Cir. 2014), held that web pages required more authentication than just a printout, but the bar remained low. The court emphasized that the proponent does not need to rule out all possibilities of forgery; they merely need to provide a "rational basis" for the jury to believe it is authentic. In the era of Photoshop, a "reasonable juror" could spot a bad edit. In the era of Diffusion, a "reasonable juror" is guessing. Studies in 2024 showed that human participants failed to identify high-quality deepfakes 48% of the time—statistically no better than a coin flip.[4] By asking jurors to "use their judgment," the courts are asking them to perform a cognitive task that is biologically impossible.

In 2024 and 2025, the Advisory Committee on Evidence Rules debated adding a specific **Rule 901(c)** to address deepfakes. The proposed rule would have shifted the burden, requiring the proponent of challenged digital evidence to prove its reliability by a "preponderance of the evidence" (a higher bar than the current "sufficiency" standard).[5]

Ultimately, the Committee rejected the amendment. Their reasoning was twofold:

pISSN:2349-5707 Volume:13,Issue 4, April -2026 eISSN:2349-5715

1. **Fear of Over-Exclusion:** They worried that a stricter rule would exclude too much genuine evidence, particularly from smartphone cameras used by bystanders.

2. **Technological Neutrality:** They argued that the rules should remain "technology neutral" and not codify standards that might become obsolete.

This article asserts that this decision was a catastrophic abdication of duty. "Technological neutrality" is a vice when the technology in question is designed to deceive. By refusing to raise the bar, the Committee effectively ruled that 1970s standards are sufficient for 2026 problems.

The consequences of this inaction are visible in the "Deepfake Defense"—a strategy where counsel suggests, without evidence, that incriminating footage is AI-generated. This is the practical application of the Liar's Dividend.

In a series of high-profile civil suits regarding autonomous vehicle crashes, defense attorneys argued that videos of the CEO making promises about "Full Self-Driving" capabilities could be deepfakes.[6] While the judge in Huang v. Tesla eventually rejected this argument, it forced the plaintiffs to spend millions on forensic analysis to prove that a public interview actually happened. This creates an economic asymmetry. A defendant can claim "deepfake" for free; the prosecution (or plaintiff) must pay $50,000 for a forensic expert to rebut it.

In the 2021 Kyle Rittenhouse trial, the defense challenged the prosecution's use of iPad "pinch-to-zoom" technology, suggesting the AI upscaling invented pixels that weren't there. In 2026, this argument has evolved. Consider a hypothetical assault case: A security camera captures the defendant punching a victim. The video is grainy. The defense argues: "Ladies and gentlemen of the jury, we know that AI can upscale video. We know AI can face-swap. This video was 'enhanced' by the police lab. How do we know the AI didn't 'hallucinate' my client's face onto the attacker?" Under the current FRE 901, the judge admits the video. But the seed of doubt is planted. If the prosecution cannot explain the "chain of custody" of the pixels (not just the file), they risk acquittal. The Deepfake Defense does not need to prove the video is fake; it only needs to prove that the video could be fake.

If "passive detection" (looking at the pixels) is failing, the law must pivot to "active authentication" (looking at the provenance). We must move from verifying the content to verifying the capture.

The C2PA Standard As of 2025, the Coalition for Content Provenance and Authenticity (C2PA) has established an open technical standard for digital provenance. Major manufacturers, including Sony (in the Alpha 9 III) and Canon (in the R1), now integrate cryptographic signing chips directly into the camera hardware.[7]

When a photo is taken, the camera cryptographically hashes the image data (including GPS, time, and pixel values) and signs it with a private key securely stored in the hardware. This creates a "manifest" that travels with the file. If a single pixel is altered by AI, the hash breaks, and the manifest indicates tampering. This is the digital equivalent of a wax seal or a DNA tamper-evident bag.

Legislative Proposal: The Verified Capture Standard (VCS) The courts must adopt a new standard that incentivizes this technology. This article proposes the adoption of a new **Federal Rule of Evidence 901(c)** and a corresponding **Rule 902(15)**.

(c) Challenged Digital Media. If a party challenges the authenticity of audio-visual evidence with a specific factual allegation of algorithmic manipulation, the proponent must demonstrate

pISSN:2349-5707 Volume:13,Issue 4, April -2026 eISSN:2349-5715

by a preponderance of the evidence that the media has not been materially altered. This burden may be satisfied by a valid cryptographic provenance record.

Draft Text of Proposed Rule 902(15): Certified Digital Provenance

(15) Certified Digital Provenance. A digital file containing a cryptographic signature from a trusted root authority, which certifies the time, location, and integrity of the capture, and which has not been invalidated by subsequent modification.

This proposal creates a "Green Lane" and a "Red Lane" for evidence.

1. Tier 1 (Green Lane): Evidence with a valid C2PA manifest is Self-Authenticating. The burden to challenge it is high; the opponent must prove the private key was stolen or the hardware hacked.

2. Tier 2 (Red Lane): Evidence without a signature (e.g., older phones, unknown sources) carries a Rebuttable Presumption of Inauthenticity if challenged. To admit it, the proponent must offer "corroborating extrinsic evidence"—witness testimony ("I saw him say that") or location metadata. The "naked eye" alone is no longer sufficient.

The United States is currently an outlier in its "laissez-faire" approach to deepfakes. Other jurisdictions have moved faster, providing models (and warnings) for US legislative reform.

China's 2023 "Provisions on the Administration of Deep Synthesis" require mandatory watermarking of all AI-generated content.[8] While effective for regulation, this "labeling" approach is insufficient for criminal evidence because criminals do not obey labeling laws. A murderer using AI to frame a rival will not politely watermark the forgery. This highlights why the US provenance approach (verifying the real) is superior to the Chinese detection approach (flagging the fake).

The EU AI Act (Article 50) mandates transparency for "high-risk" AI systems. However, European courts are also struggling with the "Deepfake Defense." In 2025, the German Federal Court of Justice (BGH) signaled that unverified social media video might require "enhanced corroboration" to support a conviction.[9] The US proposal for "Tier 2" evidence aligns with this emerging European consensus, aiming to harmonize transatlantic evidentiary standards.

1. The "Class Justice" Objection Critique: Won't the C2PA standard favor the state? Police departments can afford $5,000 Sony cameras with C2PA chips. A poor defendant filming police brutality on an old Android phone will have "Tier 2" unsigned evidence. Is this fair? Rebuttal: This is the most serious constitutional concern. A strict exclusion of unsigned media would violate the defendant's right to present a defense (Chambers v. Mississippi). Response: The proposed rule does not exclude unsigned evidence; it merely requires corroboration. If a bystander films police brutality, the video can be authenticated by the bystander's testimony ("I took this video, here is where I stood"). The rule only penalizes "orphan" evidence—video that appears on the internet with no provenance and no witness to vouch for it.

2. The "Hardware Hack" Objection Critique: What if the private key in the camera is stolen? A state actor could theoretically hack the camera's secure enclave to sign a fake video. Rebuttal: No security is perfect. However, stealing a hardware key from a Secure Enclave is infinitely harder than downloading a $5 deepfake app. The law deals in probabilities, not certainties. C2PA shifts the probability of truth from "unknown" to "highly likely." Furthermore, C2PA allows for "Revocation Lists"—if a camera key is compromised, it can be revoked globally, invalidating future signatures from that device.

pISSN:2349-5707 Volume:13,Issue 4, April -2026 eISSN:2349-5715

Conclusion. The era of "naive realism" in evidence law is over. We can no longer trust our eyes and ears. By clinging to the "reasonable juror" standard of FRE 901, the courts are engaging in a dangerous fiction—pretending that digital media in 2026 is the same as analog film in 1990.

The "Deepfake Defense" is not a future threat; it is a present reality that is eroding the fact-finding power of the jury. The "Silent Witness" is being drowned out by a cacophony of synthetic noise. The solution lies not in banning AI, but in partnering with it. By recognizing **Cryptographic Provenance** as the new gold standard, the legal system can restore the integrity of evidence.

If the Advisory Committee continues to wait, they may find that by the time they act, the public no longer believes anything they see in a courtroom. The "Liar's Dividend" will have paid out, and the cost will be the concept of reasonable doubt itself.

REFERENCES:

1. Robert Chesney & Danielle Citron, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security, 107 Calif. L. Rev. 1753, 1785 (2019).
2. Rebecca Delfino, Deepfakes on Trial: A Call to Expand the Trial Judge's Gatekeeping Role, 74 Hastings L.J. 293 (2023).
3. See United States v. Anthony, No. 24-cr-00892 (D. Mass. Mar. 15, 2025) (Order Excluding Audio Evidence).
4. Nils Köbis et al., Artificial Intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry, Computers in Human Behavior (2024).
5. Minutes of the Advisory Committee on Evidence Rules, Judicial Conference of the United States (May 15, 2025), at 4-6 (rejecting the proposed Rule 901(c)).
6. Huang v. Tesla, Inc., No. 19-cv-346663 (Cal. Super. Ct. 2024); see also Matt O'Brien, Tesla lawyers argue Musk's 'self-driving' claims could be 'deepfakes', AP News (April 27, 2023).
7. Sony Electronics, Press Release: Sony Completes Rollout of C2PA Authenticity Technology in Alpha Series, Sony Global (Oct. 30, 2025).
8. Provisions on the Administration of Deep Synthesis Internet Information Services, Cyberspace Administration of China (Jan. 10, 2023).
9. Maura R. Grossman & Paul W. Grimm, The GPT-Judge: Justice in the Age of AI, 74 Duke L.J. 1 (2024).
10. United States v. Vayner, 769 F.3d 125 (2d Cir. 2014).
11. Riana Pfefferkorn, The "Deepfake" Defense in Criminal Trials, Stanford Center for Internet and Society (2024).