

Fighting The Evidentiary Risks Of Deepfakes In Court

By **James Koukios and Tara McGrath** (July 8, 2026, 5:59 PM EDT)

Deepfake technology poses two evidentiary risks: A fabricated recording may be passed off as proof of an event that never occurred, and a genuine recording may be discredited by the mere suggestion that it could have been produced by artificial intelligence.

Courts are now confronting both risks in real cases. So far, traditional authentication rules have not been abandoned. But judges are asking more pointed questions about provenance, chain of custody, and the process by which digital evidence was collected, preserved, and presented.

This article examines how courts are developing an authentication framework for allegations that evidence is AI-generated or AI-manipulated, sometimes referred to as a deepfake. Although unsupported deepfake claims alone are unlikely to defeat admissibility, those claims can still gain traction where the exhibit's provenance is unusual or the chain of collection is weak. Proponents of digital evidence may need to offer substantially more than a witness's testimony that the evidence appears to depict what it purports to depict.

Existing Rules Still Do Most of the Work

Rule 901(a) of the Federal Rules of Evidence requires the proponent to produce sufficient evidence to support a finding that the item is what the proponent claims it is. The nonexhaustive examples in Rule 901(b) include distinctive characteristics, testimony from a witness with knowledge, and evidence describing a process or system that produces an accurate result.[1]

That flexible standard remains important in the deepfake context. A video might be authenticated by a witness who saw the event, by circumstantial details in the recording, by metadata and device information, or by forensic testimony about how the file was created and maintained.

Electronic evidence also may be self-authenticated in some circumstances under Rules 902(13) and 902(14), including through certifications addressing electronic processes or data copied from an electronic device.[2] But those tools answer different questions.

A hash value can show that a copy matches a source file, but it does not show that the source file depicts a real event. Metadata may support authenticity, but it can be incomplete, stripped, altered or ambiguous. AI-detection tools may be useful, but they are likely to raise their own questions about



James Koukios



Tara McGrath

methodology, validation, error rates and expert qualifications.

Recent Cases Chart a Middle Ground

The Iowa Court of Appeals' January decision in *State v. Amyda* demonstrates one side of the emerging line in deepfake challenges. In this case, the defendant in a sexual assault case argued that a digital video sent to the victim online might be a deepfake and that admission of a duplicate violated the best-evidence rule.[3]

The Amyda court upheld admission of the video. It emphasized the victim's testimony that she recognized her body, clothing, room, and the defendant's hand, and that those details were consistent with the only relevant time the defendant was in her room. The court concluded that this circumstantial evidence satisfied Iowa's Rule 901 analogue and that the defendant's unsupported deepfake speculation did not create a genuine question about authenticity.

The Supreme Court of Maryland's 2024 decision in *Mooney v. State*, although not a deepfake case, points in the same direction. There, the court held that video evidence may be authenticated through a combination of witness testimony and circumstantial evidence, and that a witness need not have personal knowledge of every event depicted.[4] That approach preserves the usefulness of ordinary authentication principles even as the technology surrounding digital evidence becomes more complex.

The New York Court of Appeals' February decision in *In the Matter of M.S.* illustrates the other side of the line. In a 4-3 decision, New York's high court reversed a family court's findings, which rested on video clips appearing to depict sexual abuse of a child.[5] The videos had not been found in the family home or on devices belonging to the alleged abuser but were recovered from the computer of a third party who allegedly had hacked into the family's home security cameras and who did not testify. The original camera feed was not available, a significant portion of the chain of custody was unexplained and devices recovered from the home were not offered at trial.

Applying existing authentication law, the majority in *M.S.* held that the foundation was inadequate and reversed. Although the majority did not announce a new rule, the potential for a deepfake combined with the strange origin story of the videos likely caused the majority to view the authentication requirement through a more rigorous lens. The decision noted, for example, that

the increasing prevalence of "deepfake" videos has ... rendered the method of matching circumstantial details in a video to personal observations a more suspect form of authentication; most fabricated videos "leverage" real details from real photos and videos of real places and people, then alter the pieces the person wishes to alter to create a realistic, but manipulated, video. ... In the authentication context, what matters most is whether the events depicted are as real as the proponent claims them to be, not whether there are some identifying features of the video that can be corroborated in real life.

This narrow 4-3 decision in New York underscores the close division around best practices for authenticating digital evidence in the age of AI. The dissent asserted that authentication is meant to be a low, preliminary bar: Once the proponent makes a *prima facie* showing that the evidence is what it purports to be, the evidence should be admitted. The dissent raised concerns about the blurred distinction between admissibility and weight applied in this case, imposing an unduly demanding authentication standard.

Read together, Amyda, Mooney and M.S. suggest an emerging middle ground approach to authenticating evidence that could be a deepfake. Courts are unlikely to reward unsupported deepfake claims. But when a challenger identifies concrete problems with source, custody, completeness, or forensic reliability, litigants can expect closer scrutiny of digital evidence.

Nevertheless, the decisions also highlight a concern that courts should not underestimate: the liar's dividend. If the mere existence of generative AI were enough to defeat authentication, any party could cast doubt on otherwise reliable evidence simply by raising the specter of the technology — without ever tying it to the facts of the case.

Federal Rulemaking Is Moving Slowly

The Advisory Committee on Evidence Rules, which advises the Judicial Conference of the United States, has considered, but not approved, a working draft of Rule 901(c) to address evidence alleged to be fabricated or altered by generative AI.[6]

The working draft would create a two-step framework. If enacted, the opponent first must offer evidence sufficient to support a finding that the challenged item was AI-fabricated or AI-manipulated. Only when that threshold is met would the burden shift to the proponent to establish that the item is more likely than not authentic.

This draft rule reflects the same balance emerging in the cases discussed above. It gives courts a mechanism to address serious fabrication concerns without allowing every digital exhibit to become a mini-trial about generative AI. It also distinguishes deepfake challenges from a separate proposed Rule 707, addressing machine-generated evidence.

Even so, the committee is proceeding with caution. According to its May agenda materials, the committee did not approve Rule 901(c) for publication and instead opted for further study. Those same materials cite a Federal Judicial Center survey in which just 15 of the 931 responding federal judges reported having faced a deepfake challenge. Most judges who had not yet seen such a challenge indicated they would require some initial showing before entertaining it.[7]

Guidance and Legislation Reinforce the Same Lesson

U.S. Department of Justice training materials and state and federal lawmakers are also starting to address generative AI concerns.

The March 2025 DOJ Journal of Federal Law and Practice cautioned that when AI may have been used in creating or gathering records, prosecutors should determine whether AI was used, what program was used, how the information was created, how the system was secured and what human review occurred.[8] The journal does not establish binding DOJ policy, but the article provides useful factors to consider when evidence depends on a digital process.

Similarly, in summer 2025, Louisiana amended its rules of civil procedure to require attorneys to exercise reasonable diligence to verify the authenticity of evidence before offering it and to treat undisclosed use of evidence the attorney knew or should have known was false or artificially manipulated as a violation.[9]

The DOJ training materials and Louisiana rule demonstrate that lawyers should expect judges,

regulators, and opposing counsel to ask more rigorous questions about digital provenance and understand the process before trial, not after an authentication objection.

Practical Lessons for Litigators

For proponents of evidence that could be challenged as a deepfake, authentication must begin at collection. Preserve the original device or native file when possible. Avoid relying only on screenshots, platform-compressed copies or files that have moved through unknown intermediaries. Document who collected the evidence, when, from what source, using what tool and what happened to it after collection.

Hash verification, metadata review and Rule 902 certifications should be used where they fit, but lawyers need to be precise about what they prove. Hashes are powerful evidence of integrity between copies, not proof that the underlying content is real. Metadata can corroborate authenticity but should not be oversold. In higher-risk matters, forensic expert testimony may be necessary to explain source, custody, compression, editing history, artifacts and the limits of any AI-detection analysis.

For challengers, generalized skepticism of generative AI is not enough. More effective arguments will identify specific reasons to doubt authenticity: missing originals, unexplained gaps, inconsistent metadata, evidence of editing, impossibility evidence, suspicious timing, a motive and opportunity to fabricate, or records showing use of AI tools. Discovery should target native files, source devices, platform export data, upload and download history, account logs, editing applications, render histories, prompts, and communications about the exhibit.

The evidentiary rule invoked will depend on the record. The objection may sound in Rule 901, Rule 1003, Rule 403, Rule 702, or discovery and spoliation principles. But the strategic point is the same: Build a threshold record. Courts are beginning to distinguish concrete deepfake challenges from the bare assertion that anything digital might be fake, but they also may be raising the authentication bar to account for the possibility of AI-generated evidence.

Conclusion

Deepfake technology is changing authentication practice, but it has not displaced Rule 901. Courts will continue to admit digital evidence supported by ordinary circumstantial proof when the deepfake challenge is merely speculative. At the same time, courts will demand more from proponents when the evidence has unusual provenance, missing originals, weak chain of custody, or unexplained technical gaps.

Whether you're planning to raise a deepfake challenge or respond to one, it's critical to start building the authentication strategy early. Litigators who adjust now will be better positioned when the rules eventually catch up.

James Koukios and Tara McGrath are partners at Morrison Foerster LLP.

MoFo associate Nia Anderson contributed to this article.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for

general information purposes and is not intended to be and should not be taken as legal advice.

[1] Fed. R. Evid. 901(a), (b)(1), (b)(4), (b)(9).

[2] Fed. R. Evid. 902(13), (14).

[3] State v. Amyda, No. 24-1563 (Iowa Ct. App. Jan. 28, 2026).

[4] Mooney v. State, 487 Md. 701 (2024).

[5] In the Matter of M.S. (M.H.), 2026 N.Y. Slip Op. 00825 (N.Y. Feb. 17, 2026).

[6] Advisory Committee on Evidence Rules, May 2026 Agenda Materials.

[7] Id.; Federal Judicial Center survey summarized in May 2026 Evidence Rules materials.

[8] 73 DOJ J. Fed. L. & Prac., no. 1 (2025).

[9] La. Code Civ. Proc. art. 371; 2025 La. Acts No. 250.