The Prompt

The newsletter of the ISBA's Standing Committee on Artificial Intelligence & the Practice of Law

March 2025 • Volume 1 • Number 2 •

Deepfakes in the Courtroom: Problems and Solutions

By George Bellas

The explosion of artificial intelligence (AI) has significantly impacted the practice of law. While it has improved legal research, drafting, and automating repetitive tasks, the impact of AI in the courtroom must still be confronted. The increased intrusion of AI into the legal world as a whole and the courtroom creates many challenges, both practically and ethically, in the context of litigation.

High on the list are so-called "deepfakes," a term that refers to altered or completely fabricated AI-generated images, audio, or video, that are also extremely realistic, making them difficult to discern from reality.[1] In a sense, they're AI's version of photoshopping. And the ease with which deepfakes can be created poses significant problems for courts in handling video and image evidence. We can no longer assume a recording or video is authentic when it could easily be a deepfake.

As Judge Herbert B. Dixon, Jr. of the Superior Court of the District of Columbia recently observed, "Because deepfakes are designed to gaslight the observer… any truism associated with the ancient statement 'seeing is believing' might disappear from our ethos."[2]

Deepfakes, which first appeared in 2017,[3] have been used for purposes ranging from doctored porn clips, to spoof and satire, to fraud and other crimes, as noted in a joint presentation last January by the ABA's Task Force on Law & AI, and The Bolch Judicial Institute at Duke Law School.[4] They also have appeared in the form of fictional social media accounts and voice clones. They can be created in a minute or less. We may be looking at a future in which entire movies are made using only a single scene.

In the courtroom context, deepfakes will impact evidence authenticity, witness credibility, and the integrity of the judicial process, not only because of deepfakes themselves but also because genuine evidence now can be alleged to be false, requiring this to be disproven.

Judge Dixon's article details a case in which an audio recording with the voice of a high school principal making racist and antisemitic comments about students and faculty went viral.

Ultimately, with help from two forensic analysts and a subpoena issued to Google, police traced the recording to an

email account and recovery telephone number of the school's athletic director, whose employment was pending termination.[5]

Nevertheless, "there is no foolproof way today to classify text, audio, video, or images as authentic or AI generated," wrote Professor Daniel Linna, *et al*, in the law review article, *"Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases."*[6] The authors add: "[T]hese are not challenges of a far-off future, they are already judge. Judges will increasingly need to establish best practices to deal with a potential deluge of evidentiary issues."

And Judge Dixon writes,

> "If a judge receives sworn testimony from the proponent that the evidence is a true and accurate representation of what the person said and sworn testimony from the opponent that the evidence is fake, the likely result is that the evidence will be admitted, after which the decision whether the evidence is real or fake will be left to the fact finder (judge or jury) based on the credibility of the witnesses."[7]

Among the issues confronting lawyers and judges related to AI and deepfakes:

1) Evidence Authenticity and Admissibility. Deepfakes make it difficult for courts to ascertain the authenticity of digital evidence. Traditional methods of establishing authenticity and standards of proof will be challenged. Parties may need to rely on advanced forensic tools to verify authenticity, increasing costs and complexity. And as noted by Professor Linna, et al, "Technologies designed to detect AI-generated content have proven to be unreliable, and also biased."[8]

2) Witness Credibility. Deepfakes could be used to fabricate videos or audio of individuals appearing to make incriminating or false statements, undermining their credibility. Parties may use deepfakes to intimidate witnesses by threatening to release fake yet compromising materials, discouraging them from testifying.

3) Litigation Costs. Litigants may need to hire digital forensics experts to identify and debunk deepfakes, significantly increasing litigation costs. Deepfakes can complicate the discovery process as parties may flood opposing counsel with manipulated evidence, making it even harder to discern truth from fabrication. As Professor Linna, et al, suggests, courts may have to conduct a Daubert-like hearing to establish authenticity if competing experts have different views on authenticity.[9]

4) Erosion of Trust in Evidence. Even genuine video or audio evidence may be doubted due to the potential for deepfake manipulation, leading to increased judicial skepticism and a higher burden of proof for litigants. The possibility of deepfake use may discourage pre-trial settlements, as parties could dispute the credibility of evidence.

5) Defamation and Damage Claims. In cases involving defamation or reputational damage, deepfakes can be weaponized to falsely depict individuals engaging in harmful conduct, leading to baseless but damaging claims. Demonstrating malicious intent in deepfake cases can be difficult, especially when the origin of the content is obscured.

6) Legal and Ethical Concerns. Deepfakes may be used to alter or destroy evidence intentionally, leading to allegations of spoliation and complicating the fact-finding process.

7) Impact on Discovery Rules: Courts may need to adjust discovery rules to account for the forensic challenges

posed by deepfakes, raising procedural fairness concerns.

8) Jury Challenges. Jurors lack the technical expertise to differentiate between authentic and manipulated evidence, increasing the risk of prejudicial decisions. Complex expert testimony about deepfakes can confuse jurors, making it harder for them to assess the merits of the case. The ABA/Bolch Judicial Institute presentation noted that jurors are 650% more likely to retain information provided via oral and video testimony, and that they still can be impacted despite skepticism from knowing the evidence could be fake.

9) Unregulated Use of Technology. Many jurisdictions, including Illinois and the federal courts, lack clear legal standards for addressing the creation and use of deepfakes in litigation, leaving courts to navigate uncharted territory. And, let's not ignore the problem that deepfakes often involve actors and technology across jurisdictions, complicating enforcement and accountability.

None of this is to say that AI-generated deepfakes present an insurmountable challenge for Illinois trial lawyers. But they will need to be savvy about how to confront this issue to ensure their clients get a fair hearing when opposing counsel attempts to gain an advantage by putting forth these altered or entirely fictional images, audio, or video.

Among the ways attorneys and courts can push back to reality:

Proactive Evidence Authentication. Professor Linna, et al, suggest that courts schedule an evidentiary hearing well before trial so that both sides can make their arguments about whether the evidence in question should be admitted. "[T]he judge should only admit evidence, allowing the jury to decide its disputed authenticity, after considering Rule 403 [regarding] whether its probative value is substantially outweighed by the danger of unfair prejudice to the party against whom the evidence will be used," the authors write. "Our suggested approach thus illustrates how judges can protect the integrity of jury deliberations in a manner that is consistent with the current Federal Rules of Evidence and relevant case law."[10]

A bill introduced in the California state legislature in February 2024, SB970[11] establishes standards for identifying falsified evidence. "By no later than January 1, 2026, the Judicial Council shall review the impact of artificial intelligence on the introduction of evidence in court proceedings and develop any necessary rules of court to assist courts in assessing claims that evidence that is being introduced has been generated by or manipulated by artificial intelligence."

But Judge Dixon notes that advance notice of an evidentiary issue does not necessarily solve the problem, and that if such disputes arise for the first time at trial, this "may require judges to call on their knowledge of the rules of evidence to solve the problem quickly."

Leveraging Expert Testimony. Professor Linna, et al, believe that with the rapidly improving quality of deepfakes, in the near future, nearly anyone will be able to create convincing false material, and "even experts will struggle to accurately distinguish genuine materials from fake." However, Judge Dixon's anecdote about the high school principal and athletic director suggests they will sometimes succeed.

Education for Judges and Jurors. Professor Linna and the co-authors believe judges can proactively manage evidentiary challenges related to deepfakes under the existing Federal Rules of Evidence, provided that they're sufficiently up to speed about the unique challenges this type of evidence brings with it. Mainly, this involves relevance as established in Rule 401 and authenticity under Rule 901.[12]

"This presents a low bar," they write. "If the alleged AIM is central to a matter, it will easily satisfy the relevance requirement, and satisfying the authenticity standard at this stage merely requires a show that it is more likely than not that the evidence 'is what the proponent what it is.'" Hence, the need for the proactive, pretrial conference.

Lawyers need to educate themselves and their firms on what deepfakes are and how to spot them, develop a healthy skepticism of content they encounter, and question its source. Take nothing at face value, and closely scrutinize details of that content to look for anything inconsistent with reality, such as people with more or less than five fingers.

An article published in *LegalTech News* on December 2 suggests educational resources like KnowBe4, Hook Security, and MIT Media Labs "Detect Fakes" program to get up to speed. Author Eric Hoffmaster of Innovating Computer Systems also suggests asking questions of anyone you suspect might be an AI-generated version of a given person that only the real person would know how to answer.[13]

Using AI Detection Tools. Many such tools exist to help judges and lawyers scrutinize different types of media for suspicious signs of deepfakes—or to help you confirm authenticity. Cybersecurity experts can assist the legal profession when it comes to investing in technology to deploy advanced authentication methods.

In the article, *"DeepFake-o-meter v2.0: An Open Platform for DeepFake Protection,"*[14] the authors describe the workings of the second iteration of this particular "open-source and user-friendly online platform." They write, "The platform aims to offer everyday users a convenient service for analyzing DeepFake media using multiple state-of-the-art detection algorithms. It ensures secure and private delivery of the analysis results. Furthermore, it serves as an evaluation and benchmarking platform for researchers in digital media forensics to compare the performance of multiple algorithms on the same input."

According to AIM Research, the top five tools for detecting deepfakes are: Intel's FakeCatcher, DuckDuckGoose AI, Kroop AI, TrueMedia.org, and Sensity.[15]

Changes to the Rules of Evidence. Judge Dixon's article[16] proposes three amendments to the Federal Rules of Evidence that the respective experts and commentators believe would help guide judges in handling these issues. They suggest:

- A higher standard to prove authenticity: In the law review article, *"A Break from Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes,"* the author proposes a new Rule 901(b)(11) requiring courts to go beyond a witness statement to enable the accused party to request a hearing to require the proponent to provide corroborating sources.[17]
- Judges, not juries, deciding on authenticity: In the law review article, "Deepfakes on Trial: A Call to Expand the Trial Judge's Gatekeeping Role to Protect Legal Proceedings from Technological Fakery," Professor Rebecca Delfino proposes a new Rule 901(c) based on the notion that jurors can't fairly analyze the genuineness of deepfakes. Thus, she writes, "The court must decide any question about whether the evidence is admissible," and then instruct the jury to accept the evidence as authentic and put aside generic doubts about AI if that is the judge's conclusion—while ordering opposing counsel not to exploit any such doubts.
- Placing the burden on proponents to show probative value: At the ABA Advisory Committee on Evidence Rules' meeting in April 2024, retired Judge Paul Grimm and Dr. Maura Grossman proposed a new Rule 901(c) that holds if the challenging party successfully presents evidence that challenges the authenticity of evidence as more likely than not to be a deepfake, the proponent must show that "its probative value outweighs its prejudicial effect on the party challenging the evidence," Judge Dixon writes, adding that the committee did not take action at the meeting.

In the meantime, Judge Dixon concludes "in the absence of a uniform approach in the courtroom for the admission or exclusion of audio or video evidence where there are credible arguments on both sides that the evidence is fake or authentic, the default position, unfortunately, may be to let the jury decide."

With the current state of technology, we are looking at a future in which Daubert-like hearings with competing experts analyzing the veracity of the evidence will be necessary to establish the authenticity of evidence.

Additional Resources. Recent articles have discussed this problem in greater detail than this short paper permits. Suggested reading:

Deepfakes on Trial: A Call To Expand the Trial Judge's Gatekeeping Role To Protect Legal Proceedings from Technological Fakery, 74 HASTINGS L.J. 293 (2023). The authors recognize the problem with the existing gatekeeping functions of the courts to deal with deepfakes. The courts are ill-equipped to deal with deepfakes, they believe, and the future will require lawyers to "use imagination and creativity to navigate pitfalls of proof and manage a jury's doubts and distrust about what is real."

Some state legislators are looking at enacting laws to prohibit the use of AI to falsify someone's identity or use their likeness without consent. In California, recent legislation establishes standards for identifying duped evidence in court proceedings.[18]

"Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases," Linna Jr., Daniel and Dalal, Abhishek and Gao, Chongyang and Grimm, Paul and Grossman, Maura R. and Pulice, Chiara and Subrahmanian, V.S. and Tunheim, Hon. John, Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases (August 08, 2024). Northwestern Law & Econ Research Paper No. 24-18, Northwestern Public Law Research Paper No. 24-26, Available at SSRN: **https://ssrn.com/abstract=4943841** or **http://dx.doi.org/10.2139/ssrn.4943841**

George Bellas was the chair of the initial ISBA ad hoc Artificial Intelligence Committee and a member of the Illinois Supreme Court Task Force on Artificial Intelligence.

1. See: **https://www.merriam-webster.com/dictionary/deepfake**
2. *"The 'Deepfake Defense': An Evidentiary Conumdrum,"* Judges Journal, ABA Technology Journal, June 11, 2024; see, **https://www.americanbar.org/groups/judicial/publications/judges_journal/2024/spring/deepfake-defense-evidentiary-conundrum/**
3. See:**https://www.realitydefender.com/blog/history-of-deepfakes#:~:text=The%20term%20%E2%80%9Cdeepfake%E2%80%9D%20was%20coined,open%20source%20face%2Dswapping%20technology** .
4. See: **https://www.americanbar.org/groups/centers_commissions/center-for-innovation/artificial-intelligence/ai-courts/**
5. See: **https://www.americanbar.org/groups/judicial/publications/judges_journal/2024/spring/deepfake-defense-evidentiary-conundrum/**
6. *"Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases,"* Linna Jr., Daniel and Dalal, Abhishek and Gao, Chongyang and Grimm, Paul and Grossman, Maura R. and Pulice, Chiara and Subrahmanian, V.S. and Tunheim, Hon. John, Deepfakes in Court: How Judges Can Proactively Manage Alleged AI-Generated Material in National Security Cases (August 08, 2024). Northwestern Law & Econ Research Paper No. 24-18,

Northwestern Public Law Research Paper No. 24-26, Available at SSRN: https://ssrn.com/abstract=4943841 or http://dx.doi.org/10.2139/ssrn.4943841
7. *"The 'Deepfake Defense': An Evidentiary Conumdrum,"* Judges Journal, ABA Technology Journal, June 11, 2024; see, https://www.americanbar.org/groups/judicial/publications/judges_journal/2024/spring/deepfake-defense-evidentiary-conundrum/
8. *Id.*
9. *Id.*
10. *Id.*
11. https://legiscan.com/CA/bill/SB970/2023
12. *Id.*
13. *"The Rise of AI-Generated Deepfakes: A New Cybersecurity Threat for Law Firms,"* LegalTech News, 12/2/24; See, *https://www.law.com/legaltechnews/2024/12/02/the-rise-of-ai-generated-deepfakes-a-new-cybersecurity-threat-for-law-firms/*
14. "*DeepFake-O-Meter v2.0: An Open Platform for DeepFake Detection,"* Yan Ju, Chengzhe Sun, Shan Jia, Shuwei Hou, Zhaofeng Si, Soumyya Kanti Datta, Lipeng Ke, Riky Zhou, Anita Nikolich, Siwei Lyu.
15. *"5 AI DeepFake Detector Tools for 2024,"* 6.25.24; See, *https://aimresearch.co/market-industry/5-ai-deepfake-detector-tools-for-2024?ts=1735000228*
16. *"The 'Deepfake Defense': An Evidentiary Conumdrum," supra.*
17. LaMonaga, John P. (2020) *"A Break from Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes," American University Law Review,* Vol. 69, Iss. 6, Article 5. Available at: https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?article=2221&context=aulr
18. https://www.gov.ca.gov/2024/09/17/governor-newsom-signs-bills-to-combat-deepfake-election-content/#:~:text=Advances%20in%20AI%20over%20the,democracy%2C%E2%80%9D%20said%20Assemblymember%20Berman