

Navigating the Digital Minefield: Identifying Deep Fakes and Protecting Clients in Litigation

Introduction

- **Overview:** Deep fake technology is increasingly eroding public trust in digital media, as seen in its use to fabricate speeches or create false endorsements by public figures. The rise of deep fake technology poses a significant threat to the integrity of evidence in litigation and the protection of clients in legal disputes. Attorneys must be equipped to recognize, challenge, and protect against deep fake evidence.
 - **Objective:** This paper begins with an exploration of deep fake technology and its implications, followed by real-world case examples, legal frameworks, and actionable strategies for attorneys to identify and address deep fake evidence.
-

I. Understanding Deep Fakes

1. What Are Deep Fakes?

- o Definition: Deep fakes, short for 'deep learning fakes,' rely on AI techniques such as Generative Adversarial Networks (GANs). These networks pit two AI systems against each other, one generating content and the other evaluating its authenticity, leading to highly realistic fabrications. For instance, GANs have been used to create fake celebrity interviews and voice simulations indistinguishable from authentic recordings.
- o Examples:
 - Videos altering a person's speech or actions.
 - Fabricated audio recordings of statements never made.

2. Technology Behind Deep Fakes

- o AI and machine learning models (e.g., GANs - Generative Adversarial Networks).
- o Accessible tools (e.g., DeepFaceLab, FakeApp) democratizing deep fake creation.

3. Threats Posed by Deep Fakes

- o Undermining evidence credibility.
- o Misleading witnesses, juries, and judges.
- o Damaging reputations and influencing public opinion.
- o Examples:
 - A 2019 study by Deepttrace found over 15,000 deep fake videos online, 96% of which were nonconsensual pornography.ⁱ

- Beyond personal harm, these fakes have been used in business fraud cases, such as an instance where a deep fake CEO voice was used to scam a company out of \$243,000.ⁱⁱ

II. Examples in which Deep Fake-Related cases have Emerged

1. Right of Publicity

A 2020 U.S. Court of Appeals for the Third Circuit caseⁱⁱⁱ involved artificial intelligence that seems almost indistinguishable from deepfakes. The plaintiff sued a videogame producer for the use of his likeness as one of the characters. The plaintiff's complaint alleged that the defendant had used his likeness in violation of his right of publicity.

The defendant argued that its work enjoyed First Amendment protections. The court found in favor of the defendant, reasoning that the plaintiff had been "so transformed that it has become primarily the defendant's own expression."

2. Defamation and False Light

Deepfakes have been central to lawsuits where individuals claim reputational harm due to manipulated videos or images. Courts examine whether the creation and dissemination of deepfakes constitute actionable defamation or false light invasion of privacy.

- o **Case Example:** In the U.S., defamation lawsuits based on deepfake videos have arisen where public figures argue that false and damaging portrayals were intentionally circulated. Sean "Diddy" Combs filed a lawsuit in New York City on January 22, 2025 seeking \$50 Million for defamation. He is claiming that defendants, Courtney Burgess and Ariel Mitchel, fabricated claims that have damaged his reputation. Burgess and Mitchell have produced video evidence that shows Diddy sexually assaulting celebrities and minors to NewsNation. Diddy alleges that the videos in question are fabricated and never existed.^{iv}
- o **Notable Defenses:** Public figures face a higher bar for bringing defamation claims due to the "actual malice" standard set by the Supreme Court in *New York Times Co. v. Sullivan*.^v To win a defamation suit, a public figure must show that the publisher of the defamatory statement knew "that It was false or otherwise acted with reckless disregard of whether it was false or not."^{vi}

Any attempt to regulate deepfakes must also contend with First Amendment protections. Deepfakes, particularly those involving public figures and current events, often walk a fine line between harmful misrepresentation and protected speech. Courts have long upheld the importance of free speech in cases involving offensive or outrageous statements.^{vii}

Much like defamation law, right of publicity claims require proof of harm.^{viii} Therefore, if a deepfake does not directly result in commercial gain to the poster or financial loss

to the subject,^{ix} it may be difficult for the plaintiff to demonstrate the requisite harm, especially given the fleeting nature of viral internet content and news cycles. Additionally, because right of publicity laws are primarily fashioned for commercial contexts, deepfakes which can be defended as satirical or parodical may not fall within the statutory scope.

3. Nonconsensual Pornography

Deepfakes are commonly used to create explicit content featuring individuals without their consent, leading to cases of nonconsensual pornography. Laws in some jurisdictions explicitly address this use of AI. Victims of such deepfake content have sought remedies under privacy laws, harassment statutes, or civil torts like intentional infliction of emotional distress.

- o **Case Example:** The January 2024 the publication of sexually explicit AI-generated images of Taylor Swift on X, which attracted more than 45 million views, 24,000 reposts, and hundreds of thousands of likes and bookmarks for nearly 17 hours before being shut down. Despite being eventually shut down, the images quickly spread and were reposted across accounts and social media platforms causing significant damage.
- o **Bipartisan Legislation:** The Disrupt Explicit Forged Images and non-consensual Edits Act (Defiance Act)^x of 2024 is a bipartisan legislation that aims to hold accountable those who create, share, or receive nonconsensual, sexually-explicit "deepfake" images and videos. It would allow victims to sue over such images

4. Intellectual Property

Deepfakes that utilize copyrighted materials or impersonate celebrities can lead to intellectual property disputes.

- o **Case Example:** Celebrities or companies have brought lawsuits based on unauthorized use of likeness, voice, or performance, alleging violations of right of publicity and copyright laws. On May 16, two voice actors filed a class action lawsuit against Berkeley-based LOVO, alleging that the company used AI voiceover technology to copy the actors' voices without permission. They claim it also violates their right to publicity, and the deception amounts to false advertising. The plaintiffs claimed that LOVO, that acted anonymously through a freelance marketplace, Fiverr, tricked them into providing voice samples to the company as part of a research project or for test scripts for radio ads. Lovo allegedly sampled thousands of voices to create cloned versions of voices and then sold AI voiceover subscription services to customers. LOVO also allegedly misappropriated voices, including those of A-list talent like Scarlett Johansson and Ariana Grande, to train their AI system.
- o **Relevant Laws:** U.S. copyright law (17 U.S.C. § 102) may apply if the deepfake uses protected creative works, while the right of publicity may apply to the unauthorized use of personal attributes.

5. Election Interference

Deepfakes have also been a topic in cases involving potential election interference or misleading political ads. Legal actions or regulatory discussions have centered on the harm caused by spreading misinformation through deepfake technology. The Honest Ads Act (did not pass) and state-level regulations have been introduced to prevent misuse of deepfakes in political contexts.^{xi}

- o **Example:** In 2020, a deep fake video purportedly showing Joe Biden making racially insensitive comments circulated widely on social media before being debunked. Although no lawsuits followed, the incident spurred discussions around the Honest Ads Act, aimed at regulating political ads using manipulated media. Another timely example of deepfakes portraying a public figure endorsing a candidate they never actually endorsed was a widely-shared deepfake of Taylor Swift endorsing former president Donald Trump for the 2024 presidential election.

6. Criminal Cases

Deepfake creators have been prosecuted for crimes like fraud, identity theft, or harassment:

- o **Case Example:** In a 2021 case in Pennsylvania, a woman was charged with harassment and cybercrime for creating deepfake images of her daughter's cheerleading rivals.

7. Regulatory Discussions and Precedent

While courts have not yet fully developed a robust body of precedent specifically for deepfakes, regulatory frameworks and proposed laws address their impact:

- o **Deepfake Legislation:** Several U.S. states, including California and Texas, have enacted laws that criminalize or regulate specific uses of deepfake technology, such as during elections or in cases of nonconsensual pornography.
- o **International Cases:** Lawsuits and discussions around deepfakes are also emerging in other jurisdictions, such as the EU, which emphasizes data protection and privacy laws under the General Data Protection Regulation (GDPR).

III. Legal Implications and Strategies

1. Evidentiary Challenges

- o **FRE 901:** Requirement to authenticate evidence. Under FRE 901, parties introducing evidence must authenticate it. For digital media, this involves expert testimony or forensic analysis.
- o **Case Example:** *Lorraine v. Markel American Insurance Co.*, 241 F.R.D. 534 (D. Md. 2007) - Reinforces the necessity of proper authentication for electronic evidence. "While some look to the Internet as an innovative vehicle for communication, the

Court continues to warily and wearily view it largely as one large catalyst for rumor, innuendo, and misinformation. So as to not mince words, the Court reiterates that this so-called Web provides no way of verifying the authenticity of the alleged contentions that Plaintiff wishes to rely upon...”

- o **Chain of Custody:** Ensuring a clear, documented chain of custody becomes critical to mitigate the risk of evidence tampering.
- 2. **Burden of Proof:** The existence of deepfakes may shift the burden, requiring parties to provide more robust corroboration to support the legitimacy of their evidence.
- 3. **Cross Examination Techniques:** Effective cross-examination can expose the unreliability of deepfake evidence. Attorneys should focus on identifying inconsistencies within the media, questioning the source and chain of custody, and highlighting expert testimony that challenges the authenticity of the evidence. Developing a keen eye for detail and an understanding of deepfake technology are critical skills for litigators.
- 4. **Pretrial Discovery and Motions:** proactively address potential deepfake issues during pretrial discovery. Requests for metadata, original files, and forensic analyses can help identify manipulated evidence early. Additionally, pretrial motions to exclude dubious digital evidence based on reliability and authenticity standards can mitigate the impact of deepfakes in court.
- 5. **Collaborative Efforts and Industry Standards:** Addressing the deepfake challenge requires collaboration among legal professionals, technologists, policymakers, and industry stakeholders. Establishing standardized protocols for digital evidence authentication, sharing best practices, and fostering interdisciplinary partnerships are essential for mitigating deepfake risks.
- 6. **Emerging Legislation**
 - o The Biden Administration’s Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence focused federal attention on AI in many aspects of government and society.^{xii}
 - o **California’s AB 602:** An act to add Section 1708.86 to the Civil code. Provides remedies for victims of sexually explicit deepfake content. Assembly Bill No. 602 provides individuals targeted by sexually explicit deepfake content made without their consent with a cause of action against the content's creator. Assembly Bill No. 730^{xiii} prohibits the distribution of malicious deepfake audio or visual media targeting a candidate running for public office within 60 days of an election. California’s AB 602 provides victims of nonconsensual sexually explicit deep fakes with a cause of action against creators. For example, if an individual discovers a deep fake circulating online, they can file for injunctive relief and damages, provided they demonstrate the fake’s intent to harm.

- o Examples: California’s laws prohibiting malicious use of deep fakes in elections (Cal. Civ. Code § 1798.81.5).
- o California’s Assembly Bill No. 2839, for example, was proposed in February 2024 to combat the spread of deceptive audio or visual media in the context of political campaigns.^{xiv} The bill was signed into law in September 2024, and prohibits the distribution “with malice, [of] materially deceptive content. . . .”^{xv}
- o **Texas’ SB 751:** September 1, 2019 – introduced - Criminalizes deepfakes intended to influence elections.
- o **Texas’ SB1361:** adding Sec. 21.165 – Unlawful production or distribution of sexually explicit videos to the penal code. A person commits an offense if, without the effective consent of the person appearing to be depicted, the person knowingly produces or distributes by electronic means a deepfake video that appears to depict the person with the persons intimate parts exposed or engaged in sexual conduct.
- o **Virginia’s Laws on Non-Consensual Pornography:** Address deepfakes used to harass individuals. Legislation has been enacted to make the non-consensual dissemination of sexually explicit deepfake images a Class 1 misdemeanor, addressing privacy violations and harassment. Sec. 18.2-386.1 Unlawful creation of image of another.
- o **Massachusetts:** Pending legislation aims to establish a Deepfake and Digital Provenance Task Force to protect against deepfakes used to facilitate criminal or tortious conduct.^{xvi}

IV. Identifying Deep Fakes

1. Technical Indicators

- o Visual inconsistencies: Blurring – particularly around the face, mismatched lighting, shadows that don’t align, unnatural movements.
- o Audio anomalies: Robotic tone, mismatched lip-syncing.
- o Metadata analysis: Checking creation timestamps, geolocation, and editing history.

2. Use of Forensic Tools

- o Software solutions:
 - Microsoft Video Authenticator analyzes photos and videos to detect manipulated pixels, while Deepware Scanner identifies traces of AI manipulation. However, as AI-generated media becomes more sophisticated,

even these tools face challenges, underscoring the need for ongoing advancements in forensic technology.

- Other software designed to identify manipulated media.
- o Third-party forensic analysis (e.g., Cellebrite, Magnet AXIOM).

3. Cross-Referencing Evidence

- o Corroborating content with independent sources.
- o Subpoenaing original files for analysis.
- o Employ digital watermarks to protect original content from tampering
- o Maintain a healthy skepticism toward suspicious content, especially if it seems out of character or sensationalized.
- o Example: In one case, a deep fake video allegedly showing insider trading was debunked by subpoenaing the original files, which revealed unaltered footage with timestamps corroborating the defendant's alibi.

V. Protecting Clients Against Deep Fakes

1. Preemptive Measures

- o Advising clients to maintain digital hygiene.
- o Monitoring social media and online platforms for malicious deep fakes.
- o Regularly update privacy settings on social media and other online accounts
- o Avoid oversharing – discourage sharing sensitive personal information online, which could be exploited to create deepfakes.

2. Legal Remedies

- o Injunctive relief: Immediate takedowns of harmful deep fakes.
- o Defamation and fraud lawsuits against creators or distributors.

3. Corporate Example

A global consulting firm experienced reputational harm when a deep fake video portrayed its CEO making controversial statements. By monitoring social media and employing AI detection tools, the firm was able to issue a takedown request and release a public statement refuting the video's authenticity.

4. Privacy Settings Example

Advising clients to adjust privacy settings on platforms like LinkedIn and Facebook can reduce the likelihood of their likeness being used in deep fakes. For instance, disabling public access to profile images has been shown to deter misuse.

VI. Responding to Deepfake Incidents

If a client becomes a victim of deepfake misuse, prompt action is essential. Provide the following guidance:

1. Document the Evidence

Advise clients to preserve copies of the deepfake and any related communications. Metadata and timestamps are critical for legal or forensic investigations.

2. Preservation of Evidence

- o The preservation of digital evidence (DE) presents unique problems beyond traditional evidence preservation. Digital evidence includes any information in binary form that can be useful in criminal or other legal investigations and proceedings. By its nature, digital evidence resides on physical media, but it is the content and related information, rather than the media, that are most often important. Digital evidence can exist in various formats. Physical media includes hard drives in computers and mobile devices, and external storage such as USB drives, CDs and DVDs. There are many types of physical media. Digital images and files are produced when digital evidence is copied from physical media or other systems. These may be copied from remote (cloud-based) systems, from physical systems that remain in the field or from physical systems that are seized and brought into an evidence system.
- o Other digital objects include a wide variety of digital material that does not exist as an image or file, such as an online account.
- o Law enforcement also generates potential digital evidence, especially through body worn camera (BWC) programs, in car video, and other electronic records. It consists of both a physical device and digital data.

3. Engage with Platforms

Help clients file takedown requests with social media platforms or websites hosting the deepfake. Most platforms have policies against manipulated or harmful content.

Meta's recent announcement that it will discontinue use of third-party fact-checkers on platforms like Facebook and Instagram in the United States has sparked fears of a new era of disinformation on social media. Meta is switching to a "community notes" model like that used on the social platform X.

4. Pursue Legal Remedies

Depending on the jurisdiction and nature of the deepfake, clients may have legal options, including:

- o **Defamation Lawsuits:** If the content damages their reputation.
 - o **Privacy Claims:** For the non-consensual use of their likeness.
 - o **Criminal Complaints:** If the deepfake violates criminal statutes, such as laws against cyberstalking or harassment.
-

VII. Ethical Considerations for Attorneys

1. ABA Model Rules of Professional Conduct

- o Rule 1.1: Duty of competence includes understanding emerging technologies like deep fakes.
- o Rule 3.3: Duty of candor to the tribunal requires attorneys to disclose known falsified evidence.

2. Avoiding Malpractice

- o Ensuring due diligence in verifying client-provided evidence.
- o Consulting experts when deep fake suspicion arises.

3. Case Examples

o Misuse of AI

An attorney, Steven Schwartz representing a man suing Avianca Airlines for personal injury, in a Manhattan federal court threw himself on the mercy of the court after submitting a court filing that cited at least (6) six cases that did not exist. He had used ChatGPT.^{xvii} The court held a hearing and found evidence of subjective bad faith. The judge ordered two attorneys to pay \$5,000 fines after they submitted legal briefs using bogus case citations invented by ChatGPT.^{xviii}

o Disclosing AI Use

Matter of Accounting by Weber^{xix} In New York the Surrogate's Court stated: Due to the nature of the rapid evolution of artificial intelligence and its inherent reliability issues that prior to evidence being introduced which has been generated by an artificial intelligence product or system, counsel has an affirmative duty to disclose the use of artificial intelligence and the evidence sought to be admitted should properly be subject to a *Frye* (see *Frye v. United States*, 293 F. 1013 [D.C. Cir. 1923]) hearing prior to its admission, the scope of which should be determined by the court, either in a pre-trial hearing or at the time the evidence is offered.

4. Expert witnesses

On January 10, 2025, a Stanford University Professor, Jeffery T. Hancock hired as an expert witness by the Minnesota General's Office in a lawsuit about AI-generated deepfakes. He had his testimony thrown out after it was revealed that AI itself had hallucinated citations in his court filing.^{xx} Hancock specialized in communication and AI. He admitted in testimony that he had used ChatGPT-40 to help draft his expert report. Opposing attorney's representing Representative Mary Franson and YouTuber Christopher Kohls, who were challenging the deepfake law banning AI generated deepfakes designed to influence elections. His testimony was stricken from the record, his credibility was shattered. The Attorney General requested to supplement their filings, but the judge denied the request.

VIII. Future Challenges and Opportunities

1. Advances in Technology

- o Improved deep fake detection tools.
- o AI-driven verification solutions.
- o Startups like Amber Authenticate are developing blockchain-based tools to verify media authenticity at the point of creation, providing potential solutions for mitigating deep fake threats.

2. Role of Attorneys

- o Advocating for robust legal frameworks.
 - o Educating clients and stakeholders on digital risks.
 - o Law firms can play a pivotal role in shaping policy by contributing to discussions on AI regulation, such as participating in Federal Trade Commission hearings on deceptive AI practices.
-

IX. Conclusion

• Key Takeaways:

- o Deep fakes represent a growing challenge in litigation and client protection.
- o Attorneys must combine technical tools, legal strategies, and ethical practices to combat this threat.
- o AI prioritizes coherence over accuracy – it generates responses that “sound right” rather than ensuring factual correctness.
- o AI does not “fact-check” against real legal databases unless specifically trained to do so.

- o Once AI generated errors are exposed, all expert testimony, all legal filings, and articles become suspect.
- o Verify every AI generated citation and data point.
- **Call to Action:**
 - o Encourage ongoing education on digital evidence and emerging technologies.
 - o Develop a proactive approach to deep fakes in your legal practice.
 - o Attorneys must proactively educate themselves on deep fakes, leveraging tools like forensic software and staying informed on legislative developments. By adopting a multidisciplinary approach, the legal community can address the profound challenges posed by this technology.

This CLE presentation provides attorneys with actionable insights, leveraging case law and technical expertise to navigate the challenges of deep fake evidence in litigation. Let me know if you'd like to expand on any section.

ⁱ Study titled "The State of Deepfakes" authored by Cyber security company Deeptrace.

ⁱⁱ Forbes, Damiani, Jesse 2019 September 3: A voice deepfake was used to scam a CEO out of \$243,000.00.

ⁱⁱⁱ Hamilton v. Speight, 827 Fed. Appx. 238 (3d Cir. 2020)

^{iv} Blavivt: News; Rand, Sarah, Diddy Files Defamation Lawsuit Amid Claims of Video Evidence 2025 January 24

^v 376 U.S. at 279-80.

^{vi} *Id.*

^{vii} See, e.g. *Brandenburg v. Ohio*, 395 U.S. 444 (1969) (protecting KKK speeches as free speech under the First Amendment because they were not "directed to inciting or promoting imminent lawless action..."); *Hustler Magazine v. Falwell*, 485 U.S. 46 (1988) (holding that the depiction of a well-known minister as an alcoholic and sexually immoral was protected under the First Amendment despite being offensive and "repugnant in the eyes of most[,]" even in light of potential damage to the minister's public image); *Bailey v. Iles*, 87 F.4th 275 (5th Cir. 2023) (determining that a social media post joking about police officers being instructed to shoot those infected with COVID-19 was a lawful exercise of free speech under the First Amendment).

^{viii} See, e.g. *Upper Deck Co. v. Panini Am., Inc.*, 469 F. Supp. 3d 963 (S.D. Cal. 2020); *Lemon v. Harlem Globetrotters Int'l, Inc.*, 437 F. Supp. 2d 1089 (D. Ariz. 2006).

^{ix} *Estate of Elvis Presley v. Russen*, 513 F. Supp. 1339, 1378 (D.N.J. 1981) (holding that analysis must focus on the totality of the factors bearing on whether the defendant by his activities in the marketplace has attempted to deceive or confuse the public).

^x S. 3696 – DEFANCE Act of 2024 118th Congress (2023-2024) July 23, 2024

^{xi} S. 1989, H.R. 4077 – proposed October 19, 2017 – eventually incorporated into the For the People Act and subsequently the Freedom to Vote Act – not passed.

^{xii} Exec. Order No. 14110, 88 Fed. Reg. 75191 (Nov. 1, 2023).

^{xiii} Adding Sec. 35 to the Code of Civil Procedure and repealing section 20010 of the Elections Code.

^{xiv} A.B. 2839, 2024 Assemb. (Cal. 2024).

^{xv} Cal. Elec. Code § 20012(a).

^{xvi} Bill H. 72 193rd (2023-2024).

^{xvii} Forbes, Novak, Matt 2023 May 27: Lawyer Uses ChatGPT In Federal Court And It Goes Horribly Wrong.

^{xviii} Courthouse News Service 2025, February 3; Sanctions ordered for lawyers who relied on ChatGPT artificial intelligence to prepare court brief.

^{xix} 2024 NY Slip Op 24258, Surrogate's Court, Saratoga County.

^{xx} Forbes Daniel, Lars 2025 January 29; The Irony – AI Expert's Testimony Collapses Over Fake Citations.

Authors and Presenters

This Presentation was prepared by:

Mary G. Erlingson
Erlingson Banks, PLLC
One American Place
301 Main Street, Suite 2110
Baton Rouge, LA 70801
Phone: (225) 218-4446
merlingson@erlingsonbanks.com

Woolf, McClane, Bright, Allen & Carpenter,
PLLC
Post Office Box 900
Knoxville, TN 37901
Phone: (865) 215-1043
chatmaker@wmbac.com

Jean E. Faure
Faure Holden, PC
Post Office Box 2466
1314 Central Avenue
Great Falls, Montana 59401
Phone: (406) 452-6500
jfaure@faureholden.com

Robert Draper
President | Relevant Data Technologies
www.relevantdatatech.com
6535 Shiloh Road, Suite 500
Alpharetta, GA 30005
Office: (770) 455-1445
Toll Free: (877) EDISCO-1
Cell: (678) 618-3584
robert.draper@relevantdatatech.com

Andrew J. Rutens
Galloway Wettermark Rutens, LLP
3263 Cottage Hill Road
Mobile, AL 36606
Mailing address
Post Office Box 16629
Mobile, AL 36616
Phone: (251) 476-4493
arutens@gallowayllp.com

J. Chadwick Hatmaker

