SILENT NO MORE: HOW DEEPFAKES WILL FORCE COURTS TO RECONSIDER VIDEO ADMISSION STANDARDS

Danielle C. Breen*

I. Introduction

Fake news is a term deeply embedded into our everyday vocabulary.[1] The spread of disinformation online has led to a rapid change in the way that stories are consumed and viewed by the public.[2] There are even media outlets designed exclusively for spreading false news—many under the guise of well-known networks.[3] It is more

* J.D. Candidate, Suffolk University Law School, 2021; B.A. French Language & Literature, B.A. Sociology, University of Colorado Boulder, 2016. Danielle can be reached at daniellecbreen@gmail.com.

[1] *See* Fernando Nuñez, Note, *Disinformation Legislation and Freedom of Expression*, 10 U.C. IRVINE L. REV. 783, 786 (2020) (describing fake news as information that is "routinely used to describe subjectively unfavorable content or inaccurate content that is the result of a mistake."); Holly Kathleen Hall, *Deepfake Videos: When Seeing Isn't Believing*, 27 CATH. UNIV. J. L. & TECH. 51, 53 (2018) (defining fake news as "information that is invented by people or governments for their own purposes").

[2] *See* Nuñez, *supra* note 1, at 786 (noting the speed at which false information spreads online). "Research from Massachusetts Institute of Technology (MIT) suggests that false content spreads up to six times faster than factual content on social media sites and false news stories are seventy percent more likely to be shared." *Id.*

[3] *See Don't get fooled by these fake news sites*, CBS NEWS (Nov. 17, 2019) [hereinafter *Don't get fooled*], *archived at* https://perma.cc/ARS8-7JMY (listing various fake news websites that the public should be aware of); Maxwell Library, *Evaluating Websites: Identifying Fake News Sources*, BRIDGEWATER STATE UNIV. (Nov. 17, 2019), *archived at* https://perma.cc/5E6M-GPA9 (identifying different ways to determine if a website is fake). For example, the website abcnews.com.co is designed to fool users into believing it is the authentic ABC News site. *Don't get fooled*, *supra.*

important than ever before to aptly identify disinformation.[4] A large portion of the public is skeptical of previously trusted content.[5] However, even with growing awareness of disinformation, an alarmingly high percentage of people still admit to knowingly sharing false information online.[6]

Modern technology and computer-generated imagery further complicate the ability to decipher true information.[7] Through artificial imagery known as "deepfake," it is now possible to take disinformation to the next level by creating doctored videos of events

[4] *See* Nuñez, *supra* note 1, at 785–86 (explaining the difference between disinformation and misinformation). "Disinformation is a more serious threat to freedom of expression because it is information that is deliberately created to mislead and influence the public, unlike misinformation, which may be shared under a genuine belief that its contents are truthful." *Id. See* Cat Zakrzewski, *Report urges social media companies to take down 'provably' false information*, BOS. GLOBE (Sept. 3, 2019), *archived at* https://perma.cc/5VVT-SNBQ (providing nine recommendations from a New York University report on how social media companies should combat the spread of disinformation on their platforms).

[5] *See* Joshua Benton, *Here's how much Americans trust 38 major news organizations (hint: not all that much!)*, NIEMANLAB (Oct. 5, 2018), *archived at* https://perma.cc/2P5W-RCBP (asserting that 72% of Americans trusted the media in 1972 compared to 32% in 2016). *See also* Ashley Smith-Roberts, Article, *Facebook, Fake News, and the First Amendment*, 95 DENV. L. REV. 118, 120 (2018) (describing the goal behind the spread of disinformation as "'erod[ing] trust in mainstream media, public figures, government institutions"'); Madeline Lamo & Ryan Calo, *Regulating Bot Speech*, 66 UCLA L. REV. 988, 993–94 (2019) (discussing the prevalence of bots and how quickly they can generate false content online).

[6] *See* Michael Barthel et al., *Many Americans Believe Fake News Is Sowing Confusion*, PEW RSCH. CTR. (Dec. 15, 2016), *archived at* https://perma.cc/DZB9-NN58 (providing statistics showing that 23% of Americans admitted to sharing a fake news story). *See also* Patrick Huston & Eric Bahm, *Deepfakes 2.0: The New Era of "Truth Decay"*, JUST SEC. (Apr. 14, 2020), *archived at* https://perma.cc/92WS-6Y2G (declaring that "[o]ver half of Generation Z gets its news and information primarily from social media and messaging apps on their smartphones.").

[7] *See* Rebecca Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 FORDHAM L. REV. 887, 889–90 (2019) (stating that deepfakes and technology used to create them have become "widely available"); Douglas Harris, Article, *Deepfakes: False Pornography is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 99–100 (2019) (quoting Google's CEO who believes that "artificial intelligence will change humanity more profoundly than fire."). *See also USPTO Launches Fake Specimen Informant Program*, GERBEN (Jan. 23, 2020) [hereinafter *USPTO*], *archived at* https://perma.cc/4CZF-WLZP (reporting on fake images infiltrating patent and trademark applications).

or people.[8] Deepfake creators use existing images and manipulate them to construct completely different impressions of what occurred.[9] Although some deepfakes are identifiable with the naked eye, the public's ability to recognize them is likely to decline as technological innovation enables the creation of more realistic doctored videos.[10] This new technology threatens the traditional treatment of video evidence by courts as a trustworthy representation of events.[11] The

[8] *See* Bill Hochberg, *YouTube Won't Take Down A Deepfake of Jay-Z Reading Hamlet – "To Sue Or Not To Sue"*, FORBES (May 18, 2020), *archived at* https://perma.cc/XT8S-9GZD (highlighting the issue that deepfakes pose when an individual's likeness is used without their consent); David Frum, *The Very Real Threat of Trump's Deepfake*, ATLANTIC (Apr. 17, 2020), *archived at* https://perma.cc/XA5S-9PDJ (reporting how Donald Trump intentionally tweeted an obvious deepfake video of Joe Biden while serving as president). *See also* Benjamin Goggin, *From porn to 'Game of Thrones': How deepfakes and realistic-looking fake videos hit it big*, BUS. INSIDER (July 23, 2019), *archived at* https://perma.cc/X3LB-8GP3 (outlining the long history of individuals creating deepfake videos of politicians). In April 2018, BuzzFeed created a deepfake video of Barack Obama saying things he never actually did to warn the public about the dangers of deepfake technology. *Id.* Another deepfake video in July 2018 made Alexandria Ocasio-Cortez appear as though she was unable to answer interview questions. *Id.* In May 2019, a deepfake video made Nancy Pelosi appear drunk at a political engagement. *Id.*

[9] *See* Bobby Chesney & Danielle Citron, *Deep Falls: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1756 (2019) (recounting how a deepfake video of a Parkland High School shooting survivor's speech quickly went viral).

[10] *See id.* at 1757 (noting that sophisticated deepfake technologies "are maturing rapidly."); Alex Engler, *Fighting deepfakes when detection fails*, BROOKINGS INST. (Nov. 14, 2019), *archived at* https://perma.cc/5RRB-EQ2K (describing different ways to identify deepfakes); Kevin Stankiewicz, *'Perfectly real' deepfakes will arrive in 6 months to a year, technology pioneer Hao Li says*, CNBC (Sept. 20, 2019), *archived at* https://perma.cc/GKX5-2GAG (warning that deepfake technology will soon be so realistic that humans will not be able to spot manipulated videos with the naked eye). "Carefully made deepfakes can already be very realistic, though only under certain circumstances—an attentive observer will notice that convincing deepfakes focus on individuals who don't wear glasses or have beards, and typically use a stationary camera." Engler, *supra*.

[11] *See* Jane A. Kalinski, *Jurors at the Movies: Day-In-The-Life Videos as Effective Evidentiary Tool or Unfairly Prejudicial Device?*, 27 SUFFOLK UNIV. L. REV. 789, 789–90 (1993) (illustrating the value behind video evidence as a tool for lawyers to present information in a more digestible form to juries). *See also* Jonathan Mraunac, *The Future Of Authenticating Audio And Video Evidence*, LAW360 (July 26, 2018), *archived at* https://perma.cc/VUZ2-AWGE (examining the inherent trust that juries

justice system views the ability of jurors to make determinations using visual evidence so highly that a District of Columbia Appeals Court held it constitutionally permissible for prosecutors to exclude blind persons from juries.[12] People like to believe what they see with their own eyes, and the prevalence of deepfake imagery makes this an incredibly dangerous assumption.[13]

The growing use of deepfake technology will require courts to re-evaluate the typical treatment of video evidence. A large number of jurisdictions have traditionally allowed the admission of video evidence in jury trials without a witness testifying before the jury that the video is a fair and accurate representation of what occurred. The evidence is deemed reliable behind the scenes by a judge and then presented to the jury. With deepfakes causing many people to question the authenticity of seemingly reliable videos, it is urgent that this practice changes. Deepfake videos will force courts to always require a witness to testify before the jury about the accuracy and authenticity

and society as a whole put into video evidence); David Dorfman, *Decoding Deepfakes: How do Lawyers Adapt When Seeing Isn't Always Believing?*, 80 OR. STATE BAR BULL. 18, 20 (2020) (warning that all deepfakes cannot be identified with detection technology).

[12] *See* Yael Granot et al., *In the Eyes of the Law: Perception versus Reality in Appraisals of Video Evidence*, 24 PSYCH. PUB. POL'Y & L. 93, 93 (2017) (describing the holding in United States v. Watson, 483 F.3d 828 (D.C. 2007)). Prosecutors striking blind persons from the jury demonstrates the idea that one must be able to see in order to fully comprehend all of the evidence in a case. *Id.* Visual evidence is so highly regarded in the justice system that it is seen as a way to mitigate juror bias towards other pieces of evidence in a case. *Id.*

[13] *See* Dorfman, *supra* note 11, at 20 (predicting that "in the not-so-distant future, as manipulated media becomes more prevalent, an equal concern may be the impact of deepfakes on trust in visual and audio recordings generally."); Pakinam Amer, *Deepfakes are getting better. Should we be worried?*, BOS. GLOBE (Dec. 13, 2019), *archived at* https://perma.cc/WE47-U9R7 (warning how "[v]ideo is not a substitute for truth . . . at least not anymore."); Drew Harwell, *Top AI researchers race to detect 'deepfake' videos: 'We are outgunned'*, WASH. POST (July 12, 2019), *archived at* https://perma.cc/3N8Y-C484 (discussing the dangers of highly realistic deepfakes on perception and efforts to combat them); Grace Shao, *Fake videos could be the next big problem in the 2020 elections*, CNBC (Oct. 15, 2019) [hereinafter *Deepfakes in 2020 elections*], *archived at* https://perma.cc/25MP-9AEQ (cautioning how the public will need to consider whether videos of politicians are deepfakes during the 2020 election). *See also* Philip Ewing, *What You Need To Know About Fake Video, Audio, And the 2020 Election*, NPR (Sept. 2, 2019), *archived at* https://perma.cc/S8JW-F5MX (describing how deepfakes are often detected after having a widespread negative impact on the subject's image).

of video evidence in order to mitigate jurors' doubt about the reliability of evidence.

II. History

A. The Pictorial Evidence Theory

There are traditionally two different approaches to admitting video evidence in court: the pictorial evidence theory and the silent witness theory.[14] Under the pictorial evidence theory, visual evidence is only admissible when a witness can testify before the jury that the evidence is a fair and accurate representation of what occurred.[15] This theory rests on the idea that any photographic or video evidence is just a "graphic portrayal of oral testimony," and therefore must be verified

[14] *See* Tracey Bateman Farrell, *Construction and Application of Silent Witness Theory*, 116 AM. L. REV. 1, 2 (2019) (detailing the two approaches that courts take towards admitting video evidence). *See also* FED. R. EVID. 401 (providing the test for relevant evidence); FED. R. EVID. 402 (providing circumstances under which relevant evidence is admissible); FED. R. EVID. 403 (providing circumstances under which a judge may decide to keep relevant evidence out). Under both the silent witness and pictorial evidence theory, the judge completes the 403-weighing test to determine whether to allow the visual evidence in for the jury's consideration. FED. R. EVID. 403. *See also* James Alexander Tanford, *THE PREJUDICE RULE*, IND. UNIV. MAURER SCH. L. (2014), *archived at* https://perma.cc/QYJ5-WJ6H (illustrating how the 403 balancing inquiry works). Relevant evidence will be excluded if its probative value is substantially outweighed by the risk of unfair prejudice. *Id.*

[15] *See* Farrell, *supra* note 14, at 2 (explaining the pictorial evidence standard). *See also Ex parte* Rieber, 663 So. 2d 999, 1009 (Ala. 1995) (holding that a witness's knowledge must be verified under pictorial evidence theory). The court held that the prosecution must demonstrate that the witness testifying about the events in the security tape had intimate knowledge of what occurred. *Id.* at 1011. *See also* Fisher v. State, 643 S.W.2d 571, 573 (Ark. Ct. App. 1982) (holding that videotape evidence could not be admitted without a witness under the pictorial evidence theory). The *Fisher* Court held that absent a witness verifying the events in the videotape were accurately represented and the video was not tampered with prior to trial, it could not be admitted into evidence. *Id.* at 573–74. *See also Ex parte* Fuller, 620 So. 2d 675, 679 (Ala. 1993) (providing an example of questioning that appropriately laid the foundation for the pictorial evidence theory). The prosecution questioned an investigator who was present while the Defendant gave a recorded statement. *Id.* The questions included the investigator's experience with the particular tape recorder used, whether he had played the recording before trial, and whether it was an accurate representation of what occurred. *Id.*

as correct by a witness.[16] The witness must testify that the video accurately represents the subject discussed, but they do not need to have been present at the time it was created.[17] There is also no requirement that the witness be an expert in photography or videography.[18] The witness only needs to have personal knowledge of the subject material to reliably confirm that the events presented are authentic.[19] The classic example of the pictorial evidence theory is a medical examiner testifying before the jury during a murder trial about the nature of a victim's wound.[20]

The pictorial evidence theory became increasingly relevant with the development of photo-editing technology such as Adobe, which gave parties the ability to easily alter evidence.[21] This theory

[16] *See* Jordan S. Gruber, *Foundation for Contemporaneous Videotape Evidence*, 16 AM. JURIS. PROOF FACTS 493, § 5 (2019) [hereinafter *Contemporaneous Videotape Evidence*] (setting forth why photographic and video evidence must be verified by a sponsoring witness at trial). Under the pictorial evidence theory, the "witness must testify that witness has sufficient personal knowledge of scene or events pictured or sounds recorded and that item offered accurately and reliably represents actual scene or sounds." *Id.*

[17] *See* Benjamin V. Madison III, *Scientific Evidence Symposium: Note: Seeing can be Deceiving: Photographic Evidence in a Visual Age — How much Weight does it Deserve?*, 25 WM. & MARY L. REV. 705, 708 (1984) (introducing requirements of witnesses under the pictorial evidence theory). The pictorial evidence theory is used with fingerprint and other evidence that is not meaningful to an "untrained eye" without explanation. *Id.* The pictorial evidence theory is also commonly used to depict conditions described by a witness, such as how far away something was at the time of an accident. *Id.* at 710.

[18] *See Contemporaneous Videotape Evidence*, *supra* note 16, § 5 (reiterating that the witness need only have "sufficient personal knowledge" under the pictorial evidence theory).

[19] *See id.* (maintaining that a witness may have any background so long as they have personal knowledge of events).

[20] *See* Madison, *supra* note 17, at 710 (observing how medical examiners are frequently required to provide context in murder cases because images of wounds cannot be fully understood as accurate alone). "Photographic displays allow an examiner to illustrate wounds that are difficult to conceptualize, such as numerous stab wounds, multiple bruises, or extensive damage resulting from a gunshot wound." *Id.*

[21] *See* Brian Barakat & Bronwyn Miller, *Authentication of Digital Photographs Under The "Pictorial Testimony" Theory: A Response to Critics*, 78 FLA. BAR J. 38, 38 (2004) (discussing the importance of the pictorial evidence theory in litigation). Manipulation of an image requires human action and allowing a person to testify to a jury about the content of an image prior to and after its manipulation allows the jury to still understand the image as "a true and accurate representation of what he

permits a jury to hear witness testimony regarding any alterations made to a photograph or video before concluding that the image is an accurate representation of what the witness saw.[22] This allows the introduction of video evidence that may be altered for reasons deemed permissible by a judge, such as lightening the video to better show an object.[23]

B. The Silent Witness Theory

In contrast to the pictorial evidence theory, the silent witness theory admits visual evidence absent a qualifying witness.[24] Instead, a judge deems whether there is a sufficient foundation to admit the evidence without a witness testifying before the jury.[25] The theory was

---

or she saw." *Id.* *See also* Ashley Brown, Article, *Picture ImPerfect: Photoshop Redefining Beauty in Cosmetic Advertisements, Giving False Advertising a Run For the Money*, 16 TEX. REV. ENT. & SPORTS L. 87, 90–92 (2015) (summarizing the developments of Photoshop since its creation in 1987).

[22] *See* Bakarat & Miller, *supra* note 21, at 40 (reiterating how a witness need only to testify to the foundational facts of the photograph in front of the jury for authentication). Foundational facts can be as simple as the witness being able to identify the subject in the photograph or video. *Id.* It is the responsibility of the adverse party to challenge the photo or video evidence admitted. *Id.* *See* Timothy Williams et al., *Police Body Cameras: What Do You See?*, N.Y. TIMES (Apr. 1, 2016), *archived at* https://perma.cc/JK6X-LCNA (exposing issues of perception with video evidence). *The New York Times* study showed how police body cameras only show the officer's perspective and may not be an accurate depiction of what actually occurred. *Id.* When the officer wore his chest camera, as many officers do in the United States, it appeared he was in a threatening interaction with another individual. *Id.* However, when the film of the interaction is shown from another angle, the viewer sees that the interaction was actually two people dancing together. *Id.* The reporter witnessing the two officers verified the actual interaction. *Id.*

[23] *See* Madison, *supra* note 17, at 709 (asserting how altered evidence may be admitted under the pictorial evidence theory if "the jury can understand the changes in appearance that occurred between the relevant time and the time the photograph was taken.").

[24] *See* Farrell, *supra* note 14, at 2 (indicating the components of the silent witness theory). The silent witness theory allows for admission absent a sponsoring witness because of the idea that the video evidence speaks for itself and its authenticity is corroborated. *Id.* There are no set requirements to authenticate video evidence because the facts of each case vary so widely. *Id.*

[25] *See id.* (describing how evidence is authenticated under the silent witness theory). Evidence under the silent witness theory draws its reliability from circumstances other than a witness testifying that the image or video is an authentic portrayal of

originally proposed to more easily allow X-ray images and surveillance videos into evidence.[26] The silent witness theory is a representation of the inherent trust that society places in video evidence because it demonstrates the belief that the video is a non-biased account of events.[27] Under this theory, evidence is admissible at the trial court judge's discretion upon a showing that the video was created under reliable processes and untampered with between the time it was taken and presented to the court.[28] Judges admit visual evidence under the silent witness theory as a trusted substitute for a qualifying witness's account of what happened; in other words, the process in which the evidence was obtained is deemed sufficiently reliable for admission.[29] Most jurisdictions apply the silent witness theory,

what they saw. *Id. See* Jordan S. Gruber, *Videotape Evidence*, 44 AM. JURIS. TRIALS 171, § 60 (2020) [hereinafter *Videotape Evidence*] (noting that foundational requirements for the admission of evidence are more relaxed under the silent witness theory). If there is no authenticating witness, the offering party can use "some other testimony, such as a chain-of-custody argument or 'the testimony of a photographic expert who has determined that it had not been altered in any way and was not built-up or faked,' which clearly establishes the authenticity and competency of the photographic evidence." *Id.*

[26] *See Contemporaneous Videotape Evidence*, *supra* note 16, § 5 (outlining the original justification for the adoption of the silent witness theory). X-rays do not provide any observations directly to a witness, which spurred courts to adopt the silent witness theory to have them more easily admitted into evidence. *Id. See also Videotape Evidence*, *supra* note 25, § 60 (indicating that the silent witness theory can also come into play when there is no verifying witness to verify the accuracy of the video).

[27] *See* Granot et al., *supra* note 12, at 94 (emphasizing the trust placed in video evidence by stating that viewers may be disinclined to question the creation of images and what information is excluded); Zachariah B. Parry, Note, *Digital Manipulation and Photographic Evidence: Defrauding the Courts One Thousand Words at a Time*, 2009 ILL. J. L. TECH. & POL'Y 175, 176 (2009) (commenting on the persuasive power of photographic evidence despite the fact that it "has never been easier for photos to misrepresent the truth").

[28] *See Contemporaneous Videotape Evidence*, *supra* note 16, § 5 (outlining policy considerations for using the silent witness theory). If the process behind the creation of the video is deemed inherently reliable by the judge, the evidence may "speak for itself." *Id. See* Madison, *supra* note 17, at 711 (discussing why courts apply the silent witness theory). Courts are generally reluctant to limit the use of photographic evidence, which is a large underlying policy reason behind the application of the silent witness theory. *Id.*

[29] *See Ex parte* Fuller, 620 So. 2d 675, 678 (Ala. 1993) (explaining why the silent witness theory allows evidence to be admitted absent a witness).

showcasing the insurmountable value placed in video evidence to show what actually happened in a case.[30] This assumption of reliability can create issues in the courtroom, as case law demonstrates that even multiple Supreme Court Justices have watched the same video and concluded that different versions of events occurred.[31]

Courts have adopted a variety of different approaches under the silent witness theory.[32] This is because judges emphasize the facts in each case vary so widely that it is difficult to apply one uniform standard to evaluate all video evidence.[33] Some courts set general guidelines in the application of the silent witness theory rather than mandatory standards.[34] Other courts created a standard simply

> [T]he process or mechanism substitutes for the witness's senses, and because the process or mechanism is explained before the photograph, etc., is admitted, the trust placed in its truthfulness comes from the proposition that, had a witness been there, the witness would have sensed what the photograph, etc., records.

Id. *See* Madison, *supra* note 17, at 710–11 (articulating the weight evidence is given after admission under the silent witness theory). "In practical terms, such photographic evidence assumes greater significance than photographic evidence authenticated by testimony. Instead of supplementing testimony on an issue, the photographic evidence forms an independent basis upon which the proponent may establish a fact or occurrence." *Id.* at 711.

[30] *See* Farrell, *supra* note 14, at 2 (asserting that while most jurisdictions have not expressly adopted the silent witness theory, very few have explicitly rejected it); Wagner v. State, 707 So. 2d 827, 830 (Fla. App. 1998) (setting forth the rationale behind the silent witness theory).

[31] *See* Granot et al., *supra* note 12, at 94 (reiterating what occurred in Scott v. Harris, 550 U.S. 372 (2007)). In *Scott*, the driver of a vehicle had a dashboard camera that showed him crashing. *Id.* The majority of Supreme Court Justices agreed that summary judgment was appropriate because no reasonable juror could watch the dashboard video and conclude that the driver was not reckless. *Id.* However, a few of the Justices disagreed. *Id.* *See also* Kalinski, *supra* note 11, at 798–99 (stressing the dangers of using video evidence in jury trials). Although video evidence is a powerful tool to condense tedious information into a more digestible format for jurors, it creates a strong possibility for misrepresentation and misunderstanding. *Id.*

[32] *See generally* Farrell, *supra* note 14 (listing cases in different jurisdictions outlining various approaches to the silent witness theory in civil and criminal cases).

[33] *See id.* at 2 (indicating that most courts generally have not taken a specific approach to authenticating evidence under the silent witness theory).

[34] *See id.* (introducing specific silent witness theory guidelines listed by an Indiana appeals court); State v. Anglemeyer, 691 N.W.2d 153, 161–62 (Neb. 2005) (holding that evidence is admissible when "it is a correct reproduction of what it purports to depict."); Kindred v. State, 524 N.E.2d 279, 298 (Ind. 1988) (holding that video

requiring verification of the chain of custody of the photographic or video evidence.[35] A few courts have outlined specific step-by-step processes that must be taken in order to verify video evidence.[36] One of the more restrictive step-by-step approaches is the seven-prong standard, adopted by Alabama courts, which lays out an enumerated procedure of verifying the creation process of a video before it can be admitted into evidence.[37] The goal of the seven-prong standard is to ensure that video evidence is properly deemed authentic by the court

evidence just needs a strong showing of authenticity under the facts of the case to be admitted under the silent witness theory).

[35] *See* Mraunac, *supra* note 11 (discussing the chain of custody approach to the silent witness theory). Evidence should be established that the camera was working properly, the system was reliable, and that the evidence was properly handled until the start of trial. *Id. See also* Mendoza v. Mashburn, 747 So. 2d 1159, 1172 (La. Ct. App. 1999) (reiterating that "[t]he purpose of the chain of custody rule is to assure the integrity of the evidence."); Meador v. State, 664 S.W.2d 878, 880 (Ark. Ct. App. 1984) (maintaining that "[t]he purpose of the rule requiring a chain of custody is to guard against the introduction of evidence which is not authenticated."); Nelson v. State, 687 P.2d 744, 746 (Okla. Crim. App. 1984) (asserting how "[t]he purpose of the chain-of-custody rule is to ensure that the physical evidence against the accused has not been tampered with or altered.").

[36] *See* Mraunac, *supra* note 11 (outlining the five-step test used by some jurisdictions in applying the silent witness theory). A recording's authenticity is determined by a set of five factors: "(1) evidence of time and date, (2) the presence or absence of evidence of tampering, (3) the operating condition and capability of the equipment as it relates to the accuracy and reliability of the product, (4) operating, testing and security procedures, and (5) the identification of participants depicted in the recording." *Id.*

[37] *See* Bohannon v. State, 222 So. 3d 457, 494–95 (Ala. Crim. App. 2015) (applying the seven-prong standard for the silent witness theory).

> [The seven-prong] standard requires: (1) a showing that the device or process or mechanism that produced the item being offered as evidence was capable of recording what a witness would have seen or heard had a witness been present at the scene or event recorded, (2) a showing that the operator of the device or process or mechanism was competent, (3) establishment of the authenticity and correctness of the resulting recording, photograph, videotape, etc., (4) a showing that no changes, additions, or deletions have been made, (5) a showing of the manner in which the recording, photograph, videotape, etc., was preserved, (6) identification of the speakers, or persons pictured, and (7) for criminal cases only, a showing that any statement made in the recording, tape, etc., was voluntarily made without any kind of coercion or improper inducement.

Id.

prior to it ever reaching the jury.[38] This practice assures that jurors evaluate video evidence in light of the circumstances of the case rather than question whether or not it is authentic.[39] The jurisdictional discrepancies regarding application of the silent witness theory demonstrate both the complexity of visual evidence and the value behind it.[40] In allowing this authentication process to occur outside of the courtroom, the court system reinforces the inherent trust built into video evidence by juries and society.[41] However, the silent witness theory has the potential for serious error because not all judges are familiar enough with modern photo and video editing technology to fully understand the processes through which evidence may be created—let alone to rightfully evaluate authenticity.[42]

---

[38] *See id.* at 494 (articulating the rationale behind adopting the seven-prong standard in Alabama criminal courts). *But see Videotape Evidence*, *supra* note 25, § 61 (arguing that the seven-prong test to verify chain of custody and foundational requirements is an unnecessarily high standard for the admission of video evidence).

[39] *See* Pressley v. State, 770 So. 2d 115, 132–33 (Ala. Crim. App. 1999) (providing an example of a court applying the seven-prong test to verify video evidence prior to its admission). The court held the seven-prong test was satisfied when the officer testified how the video surveillance system worked, the video was kept in his sole custody, the video was in the same condition at trial, and that there were no changes made to the video since it was in his custody. *Id. See Videotape Evidence*, *supra* note 25, § 62 (conceding that there may be a "revival of certain foundational requirements" with the prevalence of video editing).

[40] *See* Farrell, *supra* note 14, at 2 (listing different circumstances where visual evidence was admitted under the silent witness theory). "Courts have held police surveillance videotapes, bank, store, or business security surveillance videotapes, and videotapes from a bank automatic teller machine admissible . . . [o]ne court has also held that enhanced still prints made from a videotape were admissible . . . . " *Id.*

[41] *See* Patterson Dubois, *Some Observations on the Psychology of Jurors and Juries*, 53 PROC. AM. PHIL. SOC'Y 307, 316 (1914) (emphasizing the impact that visual evidence has on the perception of truth in cases); Mraunac, *supra* note 11 (highlighting common phrases used to demonstrate trust in video and photographic evidence such as "seeing is believing"). "The very fact that the photographs are handled about and continually referred to, that witnesses have seen that conditions for years . . . gradually works upon the minds of the jurors because no one can say that none of these things are so." Dubois, *supra*.

[42] *See* Melissa Whitney, *How to improve technical expertise for judges in AI-related litigation*, BROOKINGS INST. (Nov. 7, 2019), *archived at* https://perma.cc/U8BB-PQD9 (suggesting a need to have technical advisers to educate judges on technology and AI-related issues in litigation to ensure that they properly consider evidence); Judge Herbert B. Dixon Jr., *Deepfakes: More Frightening Than Photoshop on Steroids*, A.B.A. (Aug. 12, 2019), *archived at* https://perma.cc/85NZ-T2QK

C. Deepfake Technology

Deepfakes are currently the most advanced form of digital image manipulation, but they are certainly not the first.[43] Standard photoshop technology developed in 1987, and over the course of its growth has been put to increasingly deceptive use.[44] Modern examples include Adobe's Project VoCo, which makes doctored audio with a ten minute sample of the subject speaking so realistic that it is referred to as "Photoshop for the human voice."[45] Similar to photoshop, deepfakes can have harmful consequences on public perception, but

(cautioning about the challenges deepfakes will bring to the courtroom when parties have conflicting testimony about the authenticity of a video); Debra Cassens Weiss, *Should there be a duty of tech competence for judges? Survey raises questions*, A.B.A. J. (May 10, 2019) [hereinafter *Tech Competence for Judges*], *archived at* https://perma.cc/SYA7-LCDY (quoting a 2019 survey where two-thirds of judges stated that they need more e-discovery training); Riana Pfefferkorn, *'Deepfakes': A New Challenge for Trial Courts*, WASH. STATE BAR ASS'N (Mar. 13, 2019), *archived at* https://perma.cc/4B5A-KNN7 (warning how trial courts will need to become apt at confronting deepfakes).

[43] *See* Brown, *supra* note 21, at 90 (summarizing the evolution of photo editing software). Countries such as England, France, and Brazil all enacted regulation surrounding photo editing because it is so common. *Id.* at 93.

[44] *See id.* at 90 (tracing the development of photoshop back to a PhD student at the University of Michigan in 1987); Michael Scott Henderson, Note, *Applying Tort Law to Fabricated Digital Content*, 2018 UTAH L. REV. 1145, 1148–49 (2018) (discussing how courts have already had to confront edited photos in the context of child pornography); Ewing, *supra* note 13 (noting how Photoshop makes it extremely simple for the public to manipulate still images); Parry, *supra* note 27, at 182–83 (emphasizing the ease at which novice photo editors can manipulate images using Photoshop). "With a moderate amount of expertise" users can significantly alter photos to add things such as water or snow, open a door, or change what someone is wearing. Parry, *supra*.

[45] *See* Nicholas Mirra, *Putting Words in Your Mouth: The Evidentiary Impact of Emerging Voice Editing Software*, 25 RICH. J. L. & TECH. 1, 8 (2018) (explaining how VoCo creates fake audio). "The VoCo user can individually adjust each phoneme within any word in the sentence in order to create a sentence that flows as naturally as a real human statement." *Id*. The user can also alter duration and pitch of words to make the recording more realistic. *Id*. *See also* Debra Cassens Weiss, *Eckert Seamans lawyer warns about voice fakers after he nearly wired $9k to scammer*, A.B.A. J. (Mar. 10, 2020) [hereinafter *Eckert Seamans*], *archived at* https://perma.cc/QFC6-WLDE (proffering an example of an attorney falling victim to a voice-editing technology scam). A lawyer nearly wired $9,000 to a scammer after receiving a call that he believed was from his son. *Id*.

that is not the technology's sole purpose.[46] The first use of deepfakes to alter images can actually be attributed to harmless fun on social media, such as using filters to add features like dog ears to someone's face.[47] In recent years, deepfake technology has transitioned from innocent use to more problematic purposes.[48] Technological advances creating more convincing doctored photos and videos make deepfakes even more dangerous than previous digital editing technology.[49]

The mainstream term "deepfake" derives from a Reddit user's username who first began using the technology to create fake

[46] *See* Danielle S. Van Lier, *The People vs. Deepfakes: California AB 1903 Provides Criminal Charges for Deepfakes Activity to Guard Against Falsified Defaming Celebrity Online Content*, 43 L.A. LAW. 16, 18 (2020) (arguing that deepfake technology could have positive impacts on the entertainment industry by making special effects cheaper and less time-consuming); Chesney & Citron, *supra* note 9, at 1769–71 (providing examples of possible beneficial uses of deepfakes in education, art, and autonomy); Huston & Bahm, *supra* note 6 (stating that early deepfakes were "largely used for entertainment purposes."); Carlos Melendez, *It's All Fun And Games Until Someone Gets Hurt: The Implications Of Deepfakes*, FORBES (Dec. 27, 2019), *archived at* https://perma.cc/S6QN-4SPD (acknowledging several uses of deepfakes and noting that "until recently, it was harmless").

[47] *See* Melendez, *supra* note 46 (citing the origins of deepfakes in social media).

[48] *See* Charlotte Jee, *An Indian politician is using deepfake technology to win new voters*, MIT TECH. REV. (Feb. 19, 2020), *archived at* https://perma.cc/6BST-NDNZ (reporting on how the president of India's Bharatiya Janata Party appeared in the first deepfake video used for political campaigning purposes). The first video showed the Bharatiya Janata Party president speaking in English, and the second video, a deepfake, showed him speaking in Haryanvi, the Hindi dialect spoken by the political party's target voters. *Id*. *See* Melendez, *supra* note 46 (illustrating examples of recent uses of deepfake technology to portray American public figures in a negative light). An example of this is the editing of an interaction between a CNN reporter and an intern to make it appear as though the reporter attacked the intern by removing a portion of the tape where the CNN reporter said "pardon me, ma'am." *Id.*

[49] *See* Chesney & Citron, *supra* note 9, at 1771–86 (warning about the various harms deepfakes can have on individuals, organizations, and society); Grace Shao, *What 'deepfakes' are and how they may be dangerous*, CNBC (Oct. 13, 2019) [hereinafter *What Deepfakes Are*], *archived at* https://perma.cc/6T9N-5XQ4 (addressing the danger of deepfakes' accuracy deceiving people into believing falsities); Engler, *supra* note 10 (describing how "the use of this technology to manipulate facial expressions and speech, or face-swap an individual into a video, has garnered the greatest concern."). *See also* Van Lier, *supra* note 46, at 18 (warning that deepfakes can be created "in a matter of hours using an adequately powerful home computer.").

pornographic videos of celebrities.[50] It was this use of deepfake videos that originally captured public attention.[51] The word "deepfake" is derived from a combination of the phrases "deep learning" and "fake."[52] Deep learning refers to the process of training technology to become more intelligent by continuously feeding it information.[53] Deepfakes can create a false representation of events by superimposing a person's face on another's body or by changing the contents of what a person is saying.[54] Currently, deepfakes are largely used to create

[50] *See* Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 339 (2019) (referencing the beginnings of deepfake technology). An anonymous user on Reddit "superimposed images of celebrities such as Gal Gadot (*Wonder Woman*), Masie Williams (*Game of Thrones*), and Daisy Ridley (*Star Wars*) onto the bodies of adult video stars in pornographic films." *Id*. at 339–40. This led to the shocking realization that online users could create anything they wanted through the manipulation of images. *Id. See* Rachel Metz, *The number of deepfake videos online are spiking. Most are porn*, CNN (Oct. 7, 2019), *archived at* https://perma.cc/CFX3-9GWV (recounting the rise of deepfakes in 2017 on Reddit). *See also* Van Lier, *supra* note 46, at 17 (distinguishing the difference between "cheap fakes" and "deepfakes"). "Deepfakes should not be confused with the more prevalent 'shallow fakes' or 'cheap fakes,' videos manipulated through traditional video editing techniques or simply deceptively labeled to convey a narrative different from that actually depicted in the video." *Id.*

[51] *See* Hayley Duquette, Note, *Digital Fame: Amending the Right of Publicity to Combat Advances in Face-Swapping Technology*, 20 SUFFOLK UNIV. J. HIGH TECH. L. 83, 105 (2019) (detailing how the creation of celebrity deepfake pornography caught public attention). *See also* Delfino, *supra* note 7, at 893–94 (focusing on the development of "Fake App" to make video editing easier after a Reddit user first created deepfake pornography). "Before FakeApp's development, the production of realistic doctored videos was an expensive and arduous process confined to Hollywood movie studios. FakeApp's creator achieved the goal of 'mak[ing] deepfakes' technology available to people without a technical background or programming experience.'" *Id.*

[52] *See* Metz, *supra* note 50 (explaining the origin of the word deepfake).

[53] *See* Harris, *supra* note 7, at 100 (chronicling how deep learning works in the context of pornographic images).

[54] *See* Spivak, *supra* note 50, at 339 (defining deepfakes); Duquette, *supra* note 51, at 104–05 (describing how deepfakes first became controversial through superimposing celebrities' faces on pornographic actors); *What Deepfakes Are*, *supra* note 49 (refining further the definition of deepfakes). "Deepfakes refer to manipulated videos, or other digital representations produced by sophisticated artificial intelligence, that yield fabricated images and sounds that appear to be real." *What Deepfakes Are*, *supra*.

pornography, but they are also used in other avenues such as politics.[55] Examples of political deepfake videos include doctored videos of Nancy Pelosi appearing drunk at a speaking engagement, Italy's prime minster speaking in a hoarse whisper, an Indian politician speaking in a different dialect, and Barack Obama calling Donald Trump a "dipshit."[56] These examples of foreign and domestic deepfake use demonstrate the technology's ability to manipulate global public perception.[57]

To fully grasp how deepfakes work, it is helpful to understand that a video is simply a series of still images strung together in a sequence.[58] Deepfake technology uses an intelligent algorithm to

[55] *See* Metz, *supra* note 50 (offering statistics on how deepfakes are used). "While much of the coverage about deepfakes has focused on its potential to be a tool for information warfare in politics, the Deeptrace findings show the more immediate issue is porn." *Id.* Deeptrace said 96% of identified deepfakes featured pornographic content. *Id.*

[56] *See* Jee, *supra* note 48 (reporting on how an Indian political candidate used a deepfake video of him speaking in a Hindi dialect to influence a key demographic of voters); Siddharth Venkataramakrishnan, *Can you believe your eyes? How deepfakes are coming for politics*, FIN. TIMES (Oct. 24, 2019), *archived at* https://perma.cc/3VQQ-CR5T (delving into various ways deepfakes were recently used in politics to create false impressions including examples of Nancy Pelosi and the Italian Prime Minister); Kaylee Fagan, *A viral video that appeared to show Obama calling Trump a 'dips---' shows a disturbing new trend called 'deepfakes'*, BUS. INSIDER (Apr. 17, 2018), *archived at* https://perma.cc/VFD4-5HVF (explaining how a deepfake video of Barack Obama calling Donald Trump a "dipshit" went viral). Comedian Jordan Peele created the video of Barack Obama calling Donald Trump a "dipshit" in just fifty-six hours with the help of a professional video editor to demonstrate the danger behind deepfake technology. Fagan, *supra*. *See also* Sharon D. Nelson & John W. Simek, *96 Percent of Deepfake Videos Are Women Engaged in Sexual Acts*, SLAW (Mar. 25, 2020), *archived at* https://perma.cc/34ZS-LBKE (discussing the fact that most deepfakes online are pornography).

[57] *See* Duquette, *supra* note 51, at 105 (emphasizing the immorality of using someone's likeness to create deepfakes); Venkataramakrishnan, *supra* note 56 (inferring the power altered videos of politicians can have on public perception of events); Daniel Thomas, *Deepfakes: A threat to democracy or just a bit of fun?*, BBC NEWS (Jan. 23, 2020), *archived at* https://perma.cc/M643-ER3A (critiquing how politicians can use deepfakes as a defense to their behavior caught on tape). In 2018, the governor of Sao Paulo, Brazil, claimed that a video of him engaged in an orgy was a deepfake as a defense. Thomas, *supra*. The public was unable to conclusively refute his claims. *Id.*

[58] *See* Mraunac, *supra* note 11 (summarizing the basics of video evidence). Digital video cameras capture light and turn it into digital information by stringing a series

manipulate these images, otherwise referred to as deep learning.[59] Two different types of algorithms are used to create deepfake videos: discriminative and generative algorithms.[60] Discriminative algorithms classify data by looking at the subject's features and assigning it a category or label.[61] An example of this would be determining whether an email message is spam or not.[62] In contrast to discriminative algorithms, generative algorithms operate in the reverse: they first assume the category, and then determine what features make the data

of images along to create a video. *Id.* Sound waves are captured by a microphone and then turned into an electrical signal that is stored on a video tape. *Id.* Although most cameras record sound and images at the same time, it should be noted that two separate devices capture sound and images. *Id.*

[59] *See* Spivak, *supra* note 50, at 344 (articulating how deepfake technology is created through algorithm manipulation); *What Deepfakes Are*, *supra* note 49 (defining deep learning). Deepfake technology uses deep learning, which refers to algorithm arrangements that are capable of training themselves in order to make independent decisions. *What Deepfakes Are*, *supra*. Deep learning is a "subset of AI." *Id. See* Bernard Marr, *What Is Deep Learning AI? A Simple Guide With 8 Practical Examples*, FORBES (Oct. 1, 2018) [hereinafter *What is Deep Learning AI?*], *archived at* https://perma.cc/R5RN-6379 (refining further the definition of deep learning). Deep learning is when machines become capable of performing tasks that typically require human intelligence. *Id.* Deep learning occurs through an algorithm repeatedly performing tasks and slightly changing them in order to become more intelligent. *Id.* Deep learning gets its name because "the neural networks have various (deep) layers that enable learning." *Id.*

[60] *See* Spivak, *supra* note 50, at 342–43 (naming the two types of algorithms used to create deepfake videos). *See generally* Andrew Ng, Lecture, *Generative Learning Algorithms*, STAN. (2015), *archived at* https://perma.cc/DW6X-LLUJ (analyzing discriminative and generative algorithms at a high level).

[61] *See* Spivak, *supra* note 50, at 342 (explaining the function of discriminative algorithms). "'Discriminative algorithms try to classify input data; that is, given the features of a data instance, they predict a label or category to which that data belongs.'" *Id.*

[62] *See id.* (using the email spam example to demonstrate how discriminative algorithms work). *See also* Ng, *supra* note 60 (setting forth another way to think about discriminative algorithms using animal classifications).

> First, looking at elephants, we can build a model of what elephants look like. Then, looking at dogs, we can build a separate model of what dogs look like. Finally, to classify a new animal, we can match the new animal against the elephant model, and match it against the dog model, to see whether the new animal looks more like the elephants or more like the dogs we had seen in the training set.

Id.

more likely to fall into that particular group.[63] Using the same email example, if a message is identified as spam, the generative algorithm attempts to predict what features of the email increase the probability that it is considered spam.[64]

Discriminative and generative algorithms became more advanced in 2004 when University of Montreal researcher Ian Goodfellow created the Generative Adversary Network ("GAN") to produce realistic fake photos.[65] The use of GANs later expanded to encompass video editing.[66] GANs simultaneously train discriminative and generative algorithm models.[67] A neutral network within the GAN, called the generator, generates new and artificial images, while the discriminator network evaluates the authenticity of those images.[68]

---

[63] *See* Spivak, *supra* note 50, at 343 (discussing how generative algorithms work). "[A] generative model provides a way to generate data that looks like it came from the dataset. Instead of predicting a label given certain features, it attempts to predict features given a certain label." *Id.*

[64] *See id.* (applying the spam e-mail message example to generative algorithms).

[65] *See id.* (crediting Ian Goodfellow with the invention of the GAN, which "pits [discriminative and generative] algorithms against one other."); George Lawton, *Generative adversarial networks could be most powerful algorithm in AI*, TECHTARGET (June 6, 2018), *archived at* https://perma.cc/9EBM-GA63 (examining why the development of the GAN was important for artificial intelligence research). *See also* Jordan Novet, *One of Google's top A.I. people has joined Apple*, CNBC (Apr. 4, 2019), *archived at* https://perma.cc/3ZMX-JW4F (describing Ian Goodfellow's background as a widely respected researcher in artificial intelligence). The article describes Goodfellow as "the father of an AI approach known as general adversarial networks, or GANs, [whose] research is widely cited in AI literature." *Id.*

[66] *See* Spivak, *supra* note 50, at 344–45 (explaining how GANs have evolved throughout the years to encompass video editing, including infamous deepfake pornography).

[67] *See id.* at 343–45 (outlining the training processes necessary for GANs to train discriminatory and generative algorithms); Chesney & Citron, *supra* note 9, at 1760 (emphasizing how powerful GANs are). "The result far exceeds the speed, scale, and nuance of what human reviewers could achieve. Growing sophistication of the GAN approach is sure to lead to the production of increasingly convincing deep fakes." Chesney & Citron, *supra*. *See generally* IAN J. GOODFELLOW ET AL., GENERATIVE ADVERSARIAL NETS (Univ. Montreal, ed. 2014) (discussing GANs at a high technical level).

[68] *See* Chesney & Citron, *supra* note 9, at 1759 (defining a neutral network "as a kind of tabula rasa featuring a nodal network controlled by a set of numerical standards set at random"); Lawton, *supra* note 65 (asserting that "[t]he main benefit of GANs is a neural network only interprets the world through the lens of its training data");

The goal of this training is to eventually make the generator images so convincing that the discriminator network believes that they belong in the dataset.[69] The model repetition allows the GAN to become "smarter" and make more decisions on its own.[70] This is the process that creates deepfake videos.[71]

To summarize deepfake technology in the simplest way possible, consider this analogy from a professor at the University of Pennsylvania.[72] Think of the process for creating deepfakes as a personal trainer for computer software.[73] The algorithm compares images to one another to identify characteristics that it then uses to

---

Spivak, *supra* note 50, at 343–44 (explaining how a generator and discriminator network work together); Mika Westerlund, *The Emergency of Deepfake Technology: A Review*, 9 TECH. INNOVATION MGMT. REV. 39, 41 (2019) (developing further how the generator and discriminator networks interact with each other).

> These two networks called "the generator" and "the discriminator" are trained on the same dataset of images, videos, or sounds. The first then tries to create new samples that are good enough to trick the second network, which works to determine whether the new media it sees is real. That way, they drive each other to improve.

Westerlund, *supra*.

[69] *See* Spivak, *supra* note 50, at 344 (noting the point at which images become a deepfake). "After enough of this 'training,' the algorithm is refined enough to 'convincingly manipulat[e] video on the fly,' meaning it will generate images into each individual video frame such that when played regularly, the video appears seamless. This process produces a deepfake." *Id. See also* Lawton, *supra* note 65 (discussing how the training process gives GANs a better understanding of the world over time); *What is Deep Learning AI?*, *supra* note 59 (explaining how GANs can better perform and assess tasks when they are given more data to learn from).

[70] *See* Lawton, *supra* note 65 (highlighting how GANs make it possible for "'AI [to generate] previously unseen data that may be completely novel and unique, but agrees with the same type and class of data in the real world").

[71] *See* Spivak, *supra* note 50, at 344–45 (illustrating how GANs and repetitious training create deepfake videos).

> With deepfakes, the generator constructs new video frames, while the discriminator tries to discern whether the frame, with its superimposed subject, is authentic (say, an actual video frame of the original actor) or fake (a doctored video frame of the actor in a compromising position). If the discriminator cannot tell the real images from the false images, a human may not be able to either.

Id. at 345.

[72] *See Demystifying Deepfakes: 3 Truths About AI-Generated Videos*, UNIV. PA. (Nov. 14, 2019) [hereinafter *Demystifying Deepfakes*], *archived at* https://perma.cc/K8T9-TTEF (crediting professor Michael Kearns with the personal trainer deepfake analogy).

[73] *See id.* (analogizing the creation of deepfakes to personal training).

create a fake image.[74] The computer continues to identify the fake image from the real one, and each time it identifies the fake, the next false image that the computer creates is more authentic, or in "better shape" under the personal trainer analogy.[75]

D. Identifying Deepfake Videos

Before deepfakes became as sophisticated as they are today, it was fairly easy to spot a doctored video.[76] One could look for simple tells in a video suggesting that it was fake such as irregular blinking, differences in quality of certain segments, and unnatural movements.[77] More complex and modern technology is making it increasingly difficult to spot such simple changes.[78] Even if it is possible to identify alterations with the naked eye, doing so would require familiarity with the video subject's typical mannerisms in order to notice deviations that would expose the video as a deepfake.[79] The concern with the

[74] *See id.* (describing how GANs work in simpler terms).
[75] *See id.* (noting that computers train themselves on how to create deepfake videos).
[76] *See* Bernard Marr, *The Best (And Scariest) Examples Of AI-Enabled Deepfakes*, FORBES (July 22, 2019) [hereinafter *Examples of AI-Enabled Deepfakes*], *archived at* https://perma.cc/F2RR-LHEY (detailing examples of deepfakes since 2017). Some of the earliest examples of deepfakes occurred in 2017 when a Reddit user superimposed celebrities' faces on porn actors to create deepfake pornography. *Id.* These uses of deepfake technology were far less sophisticated than later uses such as the University of Washington-created deepfake editing the contents of Barack Obama's speech. *Id.*
[77] *See* Connor Levesque, *Deepfakes Explained: What, Why and How to Spot Them*, LEXLYTICS (Mar. 7, 2019), *archived at* https://perma.cc/MZL7-63WB (outlining factors to consider in determining whether a video is a deepfake); *What Deepfakes Are*, *supra* note 49 (listing ways to spot a deepfake video). Viewers can also look for tells such as a face appearing too smooth, eyes and ears not matching, and fuzziness in the video. *What Deepfakes Are*, *supra*.
[78] *See* Chesney & Citron, *supra* note 9, at 1759 (stressing the effect technological advances have on identifying deepfakes); *What Deepfakes Are*, *supra* note 49 (emphasizing the increase in access and sophistication of deepfake videos). "While the detection of doctored audio and video was once fairly straightforward, the emergence of generative technology capitalizing on machine learning promises to shift this balance." Chesney & Citron, *supra*.
[79] *See* Adrienne LaFrance, *The Technology That Will Make It Impossible for You to Believe What You See*, ATLANTIC (July 11, 2017), *archived at* https://perma.cc/2MU9-SX6R (warning how many people may not even think to question a video's authenticity if it is realistic enough). People may be unaware of

public's inability to detect deepfakes is prompting many researchers to generate deepfake identification tools, such as databases of identified deepfake videos and applications to verify digital media authenticity.[80] However, the number of researchers working on identifications tools is severely outpaced by the amount of fake videos created, with some experts estimating that the ratio of deepfake creators to researchers is as high as one hundred to one.[81]

---

the deepfake identification databases to verify a video's authenticity, and people "shar[ing] misinformation unintentionally will likely exacerbate the increasing distrust in experts who can help make sense of things." *Id.* *See* David Ingram & Jacob Ward, *How do you spot a deepfake? A clue hides within our voices, researchers say*, NBC NEWS (Dec. 14, 2019), *archived at* https://perma.cc/XT7J-VQGV (commenting on the public difficulty in spotting a deepfake video). Members of the public can be easily misled by content if are unfamiliar with the person speaking in the video. *Id.* Examples of what this may look like include Elizabeth Warren's head movements while public speaking and Barack Obama's frowning tell when he delivers bad news. *Id.*

[80] *See* Davey Alba, *Tool to Help Journalists Spot Doctored Images Is Unveiled by Jigsaw*, N.Y. TIMES (Feb. 4, 2020), *archived at* https://perma.cc/B9M5-89JS (reporting on how a new tool called "Assembler" will enable reporters to spot doctored images through "analyz[ing] an image and highlight[ing] where it thinks those [signs of manipulation] are."); Amer, *supra* note 13 (discussing research at DeepTrace Labs and the University of California to combat deepfake videos); Press Release, PR Newswire, Tamperproof Media Startup Announces Seed Round and Company Launch; Combats Digital Fraud And Deepfakes With Data Validation Platform (Mar. 18, 2020) (on file with author) [hereinafter Attestiv, Inc.] (announcing the development of Attestiv, Inc., whose "platform uses AI and blockchain technology to validate and secure the authenticity of digital media"); *What Deepfakes Are*, *supra* note 49 (describing initiatives to identify and detect deepfakes). Microsoft and Facebook are detecting and removing deepfake videos and are collaborating with universities to create a database of fake videos. *What Deepfakes Are*, *supra*. *See also* Elizabeth Culliford, *Facebook, Microsoft launch contest to detect deepfake videos*, REUTERS (Sept. 9, 2019), *archived at* https://perma.cc/9TDC-52SK (reporting on how Facebook invested $10 million in a Deepfake Detection Challenge). *See generally* SUPASORN SUWAJANAKORN ET AL., SYNTHESIZING OBAMA: LEARNING LIP SYNC FROM AUDIO (Univ. Wash., ed. 2017) (introducing the first major study conducted on deepfakes to warn the public about the technology's power to influence perception).

[81] *See* Harwell, *supra* note 13 (providing estimates on the number of researchers versus the number of deepfake creators); Amer, *supra* note 13 (describing the fight to create deepfake detection technology as a "'cat and mouse game'" with deepfake creators); Chesney & Citron, *supra* note 9, at 1762 (cautioning that "[f]or better or worse, deep-fake technology will diffuse and democratize rapidly."). *See also* Nelson & Simek, *supra* note 56 (articulating how the four biggest deepfake sites received "a combined 134 million views").

III. Premise

A. Video Evidence and Courtroom Perception

With the rising prevalence of technology in the courtroom and in crime television shows, jurors almost expect to see some form of photo or video evidence in court.[82] Even prosecutors emphasize how powerful video evidence can be in the courtroom.[83] Video evidence can visually tell the jurors what happened in a case as well as have an important emotional effect.[84] Some scholars compare the weight that jurors give to video evidence to the "CSI effect," which is the concept that the commonality of police body cameras and video surveillance leave jurors surprised—and perhaps disappointed—if there is no video evidence offered at trial.[85]

---

[82] *See* GLOBAL JUSTICE INFORMATION SHARING INITIATIVE, VIDEO EVIDENCE: A PRIMER FOR PROSECUTORS 1–2 (Bureau Just. Assistance, ed. 2016) (discussing the impact of video evidence in the courtroom). Some estimates say that video evidence is used in as many as 80% of criminal cases. *Id.* It is also noted that juries are influenced by what they see on television and expect to see video evidence more frequently as a result. *Id.*

[83] *See* John Schwartz & Katie Zezmia, *With Video Everywhere, Stark Evidence Is on Trial*, N.Y. TIMES (Dec. 8, 2010), *archived at* https://perma.cc/5GCH-F829 (reiterating how videos are typically a "main feature" of many cases). A Cornell law professor notes that video evidence is so important that during jury selection prosecutors often consider how potential jurors will able to process video evidence in highly emotional cases. *Id.*

[84] *See id.* (contrasting different opinions on the use of video evidence in court and its impact). Legal experts say the rise of accessible technology such as cell phone camera will force judges to reevaluate the way they admit video evidence. *Id.* The District Attorney in Suffolk County, Massachusetts said it is "a powerful tool for us in determining the truth." *Id.* A defense attorney noted that the power of video evidence requires defense attorneys to regularly file motions to challenge video evidence. *Id.* A juror noted in a sexual molestation case that the video was not necessary to convict but still had an intense emotional impact on him while considering the case. *Id.*

[85] *See* Paula Hannaford-Agor, *Are Body-Worn Cameras the New CSI Effect?*, 30 CT. MANAGER 3 (2015) (inferring that jurors expect to see some form of video evidence due to the growing prevalence of surveillance and cell phone video in society). *See also* Honorable Donald E. Shelton, *Juror Expectations for Scientific Evidence in Criminal Cases: Perceptions and Reality about the "CSI Effect" Myth*, 27 W. MICH. UNIV. COOLEY L. REV. 1 (2010) (noting that jurors who watch CSI have higher expectations of seeing some form of forensic evidence in criminal cases). A study

Disagreements can arise over what occurred even with untampered video evidence.[86] A study conducted by *The New York Times* explored perception disputes through reporters demonstrating different points of view of the same incident by filming a staged interaction from different angles.[87] They compared the footage taken by two police officers' body cameras with the footage taken at the same time by a reporter at a different angle.[88] Each of the officers' body cameras made it appear as though they had an altercation, whereas it was evident from the reporter's camera angle that the officers were dancing with each other.[89] Another illustration of the complexities of perception was seen in *Scott v. Harris*, where dashboard video evidence played a large role in the United States Supreme Court's holding that a police officer used unjust force in a car chase and caused the plaintiff's injuries.[90] In contrast to the majority's perception of the dashboard video, Justice Stevens and several lower court judges disagreed and posited a different version of what they believe occurred.[91]

---

revealed 58.3% of jurors expect to see some kind of scientific evidence in every criminal case, 42.1% expect DNA evidence in every criminal case, and 56.5% of jurors expect to see fingerprint evidence in every criminal case. *Id.*

[86] *See* Williams et al., *supra* note 22 (showing disagreement over what occurred in a police body camera video). *See also* Granot et al., *supra* note 12, at 96 (describing a study where participant viewers missed crucial details when asked to focus on a specific portion of a video). In one study, 46% of people watching a scene with people passing a ball failed to see distractions coming right through the middle of the video. *Id.*

[87] *See* Williams et al., *supra* note 22 (explaining the different camera angles used in the study). One angle was filmed from a police officer's body camera and the other was filmed from a reporter's vantage point where both parties could be seen. *Id.*

[88] *See id.* (setting forth how the footage of the incident was compared in the study).

[89] *See id.* (highlighting the bias that comes from how video footage is obtained and presented).

> In this case, the "struggle" appears to be far more involved than it actually is because the camera is mounted on the officer's chest, producing herky-jerky movements that exaggerate what's going on. Even if the camera was on the officer's glasses or hat, the up-close footage would be confusing — proof that perspective matters.

Id.

[90] *See* Scott v. Harris, 550 U.S. 372, 375–76 (2007) (providing the facts of the case).

[91] *See id.* at 378–81 (detailing the Supreme Court's view of what occurred on the videotape evidence). The Court rejects the lower court's interpretation that the Respondent was driving with care, holding that it clearly appeared that Respondent was fleeing from the police. *Id. See* Granot et al., *supra* note 12, at 94 (emphasizing

B. Disinformation and its Consequences on Public Perception

Further complicating the impression of events is the alarming prevalence of disinformation in the era of fake news.[92] In the United States, most people are familiar with, or at the very least have heard of, the impact of false information on the 2016 presidential election.[93] To illustrate the effect of disinformation, consider the fact that a false story claiming the Pope endorsed Donald Trump received over two million Facebook engagements.[94] In contrast, the top *New York Times* news story that year received just over 300,000 online engagements, highlighting the power of a sensational false headline in grabbing public attention.[95] Many news commentators suggest that Donald Trump won the 2016 presidential election because of the prevalence

the fact that some Americans had widely different interpretations of the video that the Supreme Court Justices used to make their judgment). A quarter of 1,300 Americans surveyed disagreed with the Supreme Court Justices' views on what occurred in the video. *Id.*

[92] *See* Nuñez, *supra* note 1, at 785 (reiterating the difference between misinformation and disinformation is that disinformation "is deliberately created to mislead and influence the public"); Barthel et al., *supra* note 6 (providing statistics on the prevalence of sharing fake news). 23% of Americans have shared a fake news story, knowingly or not, and 64% say that fake news stories cause confusion about current events. Barthel et al., *supra*. *See also* Hall, *supra* note 1, at 55 (sharing an MIT Twitter study that found a false story spread "six times more rapidly than a true story").

[93] *See* Danielle Kurtzleben, *Did Fake News On Facebook Help Elect Trump? Here's What We Know*, NPR (Apr. 11, 2018), *archived at* https://perma.cc/PTX9-CMC4 (discussing the various impacts of fake news on the 2016 presidential election). More than a quarter of voters in the 2016 election visited a fake news site. *Id.*

[94] *See* Hannah Ritchie, *Read all about it: The biggest fake news stories of 2016*, CNBC (Dec. 30, 2016), *archived at* https://perma.cc/8MTB-7W8U (recapping the top fake news stories of the 2016 presidential election).

[95] *See* Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 J. ECON. PERSP. 211, 212 (2017) (focusing on the fact that popular fake news stories were shared more often on Facebook than popular mainstream news stories); Lamo & Calo, *supra* note 5, at 999 (asserting that the prevalence of disinformation is furthered through bots which "can help spread false or misleading news or else stoke national strife during a crisis or other salient news event."); Nelson & Simek, *supra* note 56 (reporting that there is "no absence of demand for these images" as there were 134 million views of deepfake videos on the four most popular deepfake sites).

of disinformation, demonstrating its dangerous effects on the public.[96] Coupled with the disconcerting rise of false news stories is the expanding presence of deepfake videos, which are projected to increase in the coming years.[97] Deepfakes are gaining so much traction that companies are forming specifically to create doctored videos for customers.[98] Like the effect that false news stories had on the 2016 presidential election, deepfake videos similarly can have a massive impact on public perception of events.[99] Just as people are

[96] *See* Lamo & Calo, *supra* note 5, at 998 (citing statistics showing that bot activity spreading stories on election day was "five-to-one pro-Trump to pro-Clinton"); Allcott & Gentzkow, *supra* note 95, at 212 (stating commentators suggested that Donald Trump won the 2016 election because of the prevalence of fake news stories); Caitlin Dewey, *Facebook fake-news writer: 'I think Donald Trump is in the White House because of me'*, WASH. POST (Nov. 17, 2019), *archived at* https://perma.cc/S3HW-Y3NQ (interviewing a fake news reporter who stresses that many people take stories at face value); Hannah Jane Parkinson, *Click and elect: how fake news helped Donald Trump win a real election*, GUARDIAN (Nov. 14, 2016), *archived at* https://perma.cc/9Q8Q-9YBF (emphasizing that millions of people clicked on political fake news stories shared on Facebook); Max Read, *Donald Trump Won Because of Facebook*, N.Y. MAG. (Nov. 9, 2016), *archived at* https://perma.cc/DGY8-9Z5K (focusing on the alarming fact that many online users simply read fake news headlines and discuss them with others as though they are true).

[97] *See* Metz, *supra* note 50 (noting the huge increase in the number of deepfake videos online counted by a study in October 2019). Deeptrace said there was an 84% increase in deepfake videos online since its last count in December 2018. *Id. See also* Harwell, *supra* note 13 (comparing the number of researchers creating deepfake detection technology to the number of people creating deepfake videos). "'We are outgunned,' said Hany Farid, a computer-science professor and digital-forensics expert at the University of California at Berkeley. 'The number of people working on the video-synthesis side, as opposed to the detector side, is 100 to 1.'" *Id.*

[98] *See* Delfino, *supra* note 7, at 893 (assessing the impact of "Fake App" by stating "[b]efore [its] development, the production of realistic doctored videos was an expensive and arduous process"); Chesney & Citron, *supra* note 9, at 1763 (recounting how quickly the public began using "Fake App" to manipulate images once it appeared online); Michael Andor Brodeur, *The future of your face on the Internet*, BOS. GLOBE (Feb. 13, 2020), *archived at* https://perma.cc/55EC-2FP3 (reporting on how "[o]ur face can be peeled off and deepfaked onto other bodies" through apps Morphin and Doublicat); Nelson & Simek, *supra* note 56 (discussing the easy accessibility to deepfake creation technology through avenues such as the $50 app called DeepNude); Venkataramakrishnan, *supra* note 56 (providing that Deepfakes Web charges two dollars an hour to create deepfake videos).

[99] *See* Amer, *supra* note 13 (noting that while "one single incident of a deepfake may not lead to a permanent distortion of facts" multiple instances will cause public distrust in photos and videos); Huston & Bahm, *supra* note 6 (stating that "deepfakes

likely to believe what they read online is true, studies have shown over and over again that people tend to believe what they see, despite knowing that videos can misrepresent facts.[100] Even judges can disagree over legal implications of video evidence to the point where cases are decided differently on appeal because of conflicting evidentiary interpretations.[101]

C. Legislative Action to Curb Deepfakes

In December 2019, the President signed the first bill concerning deepfakes into law.[102] The law, a part of the National Defense Authorization Act for Fiscal Year 2020, requires a report on foreign use of deepfakes, notification to Congress of foreign deepfake misuse to influence elections, and the creation of a competition to encourage deepfake research.[103] A number of other bills regarding

pose 'an extraordinary threat to the sound functioning of government, foundations of commerce and social fabric.'"); Venkataramakrishnan, *supra* note 56 (citing examples of deepfakes used against powerful public figures such as Matteo Renzi, Italy's former prime minister, and Speaker of the House, Nancy Pelosi); Fagan, *supra* note 56 (discussing how a deepfake video of Barack Obama calling Donald Trump a "dipshit" quickly became widely circulated).

[100] *See* Granot et al., *supra* note 12, at 97–98 (warning how powerful video evidence can be in convincing people that a fake event occurred). In a study conducted by a bank where no participants illicitly took money, the bank was still able to convince participants that they stole money after showing them a doctored video in which it appeared that they stole. *Id.* After watching the video, while knowing that they did not steal, participants would confess to taking money from the bank. *Id.*

[101] *See id.* at 97 (summarizing McDowell v. Sherrer, 374 Fed. Appx. 288 (3d. Cir. 2010) where judges disagreed on what was demonstrated through video evidence). In *McDowell*, a district court judge granted summary judgment for a prison guard after viewing a video of a prison altercation and finding it conflicted with the prisoner's claim that the guard used excessive force against him. *Id.* An appeals court reversed the grant of summary judgment, holding that the district court judge assigned too much weight to what she believed she saw in the video and should have left the question of excessive force to the jury. *Id.*

[102] *See* Matthew Ferraro et al., *First Federal Legislation on Deepfakes Signed Into Law*, WILMERHALE (Dec. 23, 2019), *archived at* https://perma.cc/C9CV-FQFG (announcing that Donald Trump signed the first law concerning deepfakes on December 20, 2019).

[103] *See id.* (summarizing the report that will be delivered to Congress in 2020 on deepfakes by the Director of National Intelligence). The report to Congress is required to include (1) the technological capabilities of foreign countries to create

deepfake reporting and research are also pending in Congress.[104] The most significant is H.R. 3230, otherwise known as the Deepfakes Accountability Act, which mandates transparency when videos are falsely edited in order to combat the spread of disinformation and incorrect perception.[105] The Deepfakes Accountability Act calls for specific ratifications for tampered videos, such as mandatory inclusion of digital watermarks and express disclosures that describe how the video was altered.[106]

deepfakes; (2) how disinformation from foreign governments could harm the United States' elections; (3) what technology the United States can develop to combat deepfake attacks; (4) current deepfake capabilities of the United States; (5) an explanation of what is currently being done regarding deepfakes in the United States; and (6) recommendations for additional needs to combat deepfakes. *Id. See* National Defense Authorization Act for Fiscal Year 2020, S. 1790, 116th Cong. (2019) (enacted) (authorizing the appropriations and policies for the Department of Defense).

[104] *See* Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R. 3230, 116th Cong. (2019) (providing guidelines to regulating and marking tampered videos); Deepfake Report Act of 2019, S. 2065, 116th Cong. (2019) (requiring the Department of Homeland Security to report on state of "digital content forgery technology" during specified periods); Identifying Outputs of Generative Adversarial Networks Act, H.R. 4355, 116th Cong. (2019) (asking for federal support in manipulated media research).

[105] *See* H.R. 3230 (naming proposed legislation by Congress that protects the public from disinformation spread through deepfakes); Dorfman, *supra* note 11, at 21 (asserting that the Deepfakes Accountability Act is "[t]he most significant bill in Congress"). *See also* Chesney & Citron, *supra* note 9, at 1758 (highlighting the issues of subject consent that arise with the creation of deepfakes). "Although deep fakes can be created with the consent of people being featured, more often they will be created without it." *Id.*

[106] *See* H.R. 3230 § 1041 (a)–(e) (specifying ratifications needed for altered videos under the Texas law). The new regulations include providing a digital watermark on the altered image; and an audiovisual disclosure; a visual disclosure; or an audio-disclosure. *Id. See also* Daniel Lipkowitz, Article, *Manipulated Realty, Menaced Democracy: An Assessment of the Deep Fakes Accountability Act of 2019*, 2020 N.Y.U. J. LEGIS. & PUB. POL'Y QUORUM 30, 31 (2020) (critiquing the reforms outlined in the DeepFakes Accountability Act). Watermarks can be easily removed and it is difficult to find the creators of false content. *Id.* However, the legislation is a step in the right direction towards regulating deepfakes because it (1) draws a clear line between criminal and non-criminal deepfakes; and (2) current criminal and tort law does not adequately address harms caused by deepfakes. *Id.*

However, there is still an alarming lack of state legislation pertaining to deepfake video use.[107] This is incredibly dangerous because accurate public perception of events is extremely vulnerable when powerful individuals and organizations endorse deepfake videos.[108] Despite the shortcomings of most states, others realized the imminent threat of deepfakes and took action.[109] Virginia amended its law banning the use of images "with the intent to coerce, harass, or intimidate" another person to encompass falsely created videos in early 2019.[110] Texas then became the first state to amend its Election Code to explicitly criminalize the creation and distribution of deepfakes

[107] *See* Kenneth Artz, *Texas Outlaws 'Deepfakes'—but the Legal System May Not Be Able to Stop Them*, LAW.COM (Oct. 11, 2019), *archived at* https://perma.cc/VDM3-T444 (noting the lack of action in the legal system to criminalize deepfake videos); Dorfman, *supra* note 11, at 21 (foreshadowing that more states will begin to enact deepfake legislation when they realize deepfakes can "manipulate financial markets, slander professional and personal rivals, incite violence and blackmail people falsely depicted as engaging in unethical conduct.").

[108] *See* Venkataramakrishnan, *supra* note 56 (cautioning on the immense power deepfakes have when shared by public officials). Donald Trump shared an altered video of Nancy Pelosi on Twitter which received 30,000 retweets and 90,000 likes. *Id.* *See* Lauren Feiner, *Facebook says the doctored Nancy Pelosi video used to question her mental state and viewed millions of times will stay up*, CNBC (May 24, 2019), *archived at* https://perma.cc/ZGZ2-G6S4 (emphasizing the impact of the deepfake video of Nancy Pelosi on public perception). Rudy Giuliani, Donald Trump's attorney at the time, shared the false video and stated "[w]hat is wrong with Nancy Pelosi? Her speech pattern is bizarre." *Id.* The video was slowed down by 75% to make it appear that Pelosi was slurring her words. *Id.*

[109] *See* Lucas Ropek, *Handful of States Begin Legislating "Deepfake" Videos*, GOV'T TECH. (Apr. 30, 2019), *archived at* https://perma.cc/WFG4-9WGZ (reporting on some states' efforts to create legislation criminalizing deepfakes). The Pentagon announced an effort to research ways to combat false videos and several states began considering federal legislation. *Id.* These efforts are largely based out of concern for false information influencing elections. *Id.* *See also* Dorfman, *supra* note 11, at 21 (predicting that states banning deep fakes altogether rather than labeling them may face constitutionality issues in the future).

[110] *See* VA. CODE ANN. § 18.2–386.2 (2020) (criminalizing falsely created pornographic images under Virginia law); H.B. 2678, 2019–2020 Leg., Reg. Sess. (Va. 2019) (providing the amended language to Virginia's original law criminalizing the malicious distribution of pornographic images without the subject's consent); Robert Volker & Henry Ajder, *Analyzing the Commodization of Deepfakes*, 2020 N.Y.U. J. LEGIS. & PUB. POL'Y QUORUM 22, 27 (2020) (crediting Virginia as the first state to criminalize "nonconsensual, 'falsely created,' explicit images and videos . . . a Class 1 misdemeanor.").

intended to harm a political candidate in September 2019.[111] Under the Texas law, it is a criminal offense to knowingly post a manipulated video of a political candidate within thirty days prior to an election.[112] Shortly after Texas, California passed a similar election bill in October 2019.[113] The California legislation makes it illegal for anyone to knowingly post a deepfake video relating to a political candidate within sixty days prior to an election.[114] California also amended its Penal Code in January 2020 to specifically define deepfakes and outline criminal penalties for the creation of malicious deepfake videos.[115]

---

[111] *See* Matthew Ferraro, *Texas Law Could Signal More State, Federal Deepfake Bans*, LAW360 (Sept. 6, 2019) [hereinafter *Texas Deepfake Law*], *archived at* https://perma.cc/6N2D-T6HY (observing that Texas is first state to enact deepfake legislation criminalizing deepfake-related conduct). Texas is the first state to enact legislation banning the creation of deepfake videos and the second state to enact criminal penalties for the distribution of deepfake videos. *Id. See also* Volker & Ajder, *supra* note 110, at 27 (noting that Virginia was the first state to criminalize "'falsely' created explicit images and videos," although the legislation did not explicitly call them deepfakes).

[112] *See* 15 TEX. ELEC. CODE § 255.004(e) (2019) (defining the term deepfake in the Texas law). "[A] video created" with artificial intelligence that, with the intent to deceive, appears to depict a real person performing an action that did not occur in reality." *Id.*

[113] *See* Will Fischer, *California's governor signed new deepfake laws for politics and porn, but experts say they threaten free speech*, BUS. INSIDER (Oct. 10. 2019), *archived at* https://perma.cc/D2QQ-HF5F (describing the California deepfake legislation on elections and reactions to it). Critics of the deepfake legislation say it may hurt free speech principles under the First Amendment because "[t]he law is overbroad, vague, and subjective." *Id.* Assemblyman Bernman countered this idea by stating, "'[w]hile the First Amendment gives you the right to say whatever you want, it does not give you the right to put your words into my mouth, or to use AI technology to take my body and make it look like I did something I never did.'" *Id.*

[114] *See* Assemb. B. 730, 2019–2020 Leg., Reg. Sess. (Cal. 2019) (setting forth how both California's Code of Civil Procedure and Elections Code were amended through the legislation). The bill's protections are active until January 1, 2023. *Id.*

[115] *See* Assemb. B. 1903, 2019–2020 Leg., Reg. Sess. (Cal. 2020) (defining deepfakes and acts using them which are criminalized under California law).

> [A deepfake is] a recording that has been created or altered in a manner that it would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of the individual depicted in the recording . . . [the] bill would also criminally prohibit a person from preparing, producing, or developing, without the depicted individual's consent, a deepfake depicting sexual conduct.

D. Deepfake Evidence in the Courtroom

In response to growing public awareness and newly introduced deepfake legislation, legal scholars are finally considering how to handle deepfake videos in the courtroom.[116] Due to the nature of deepfakes and their modernity, there is hardly any case law suggesting how to confront them in court.[117] However, one case—*United States v. Chapman*—provides helpful language from an expert witness authenticating video evidence in court which may be useful to consider in the context of deepfakes.[118] In *Chapman*, an undercover police officer filmed the defendant engaging in drug activity, and at trial the defense offered two witnesses to establish whether any video

Id. *See also* Van Lier, *supra* note 46, at 21 (praising the California legislature for enacting a criminal deepfake legislation in January 2020).

[116] *See* Mraunac, *supra* note 11 (explaining the need for the legal industry to respond to fake video evidence) (alteration in original); Huston & Bahm, *supra* note 6 (warning that "[c]riminals could use deep fakes to defraud victims, manipulate markets, and submit false evidence to courts."). "The legal industry, along with other sectors, will have to take serious action in order to avoid litigating and living in a world where audio and video recordings are presumed untrustworthy." Mraunac, *supra*. *See also* Catherine Stupp, *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case*, WALL ST. J. (Aug. 30, 2019), *archived at* https://perma.cc/5CM9-YD92 (providing a high-profile criminal example of a deepfake scam). A CEO fell victim to believing he spoke with his boss about a money transfer and mistakenly sent funds to a scammer using deepfake technology to sound like his boss. *Id.* This is one example of how even those who work closely to an individual can fall victim to deepfakes, which draws doubt on how those removed from a situation may be able to identify a fake video or audio recording. *Id.*

[117] *See* Mraunac, *supra* note 11 (noting a lack of formal guidance on confronting deepfakes in case law). Although courts are obviously concerned with the reliability of video evidence, there is little case law actually dealing with manipulated videos and whether they should be admitted into evidence because of their relative novelty in the last few years. *Id.*

[118] *See* U.S. v. Chapman, 804 F.3d 895, 897–98 (7th Cir. 2015) (providing the factual background of the case). An officer acted as an undercover wearing a Hawk recording device while engaging with the defendant in several drug transactions. *Id.* at 897. Each time the officer began recording, he stated his name and the time. *Id.* at 897–98. If there were any gaps in the recording, it was shown in the device's time stamp. *Id.* at 898. After the device was deactivated, the agent returned it to the FBI office where the data was downloaded onto a DVD through an FBI software program. *Id.*

tampering occurred.[119] This expert testimony regarding the verification of the video evidence would fall under the pictorial evidence theory, discussed in Part II of this Note, which requires that a witness testify as to the video's authenticity.[120] The expert witnesses verified several factors, including that the video was consistent with FBI procedures, that the data collection time was verified, and that the individual files were verified.[121] This application of the pictorial evidence theory allowed the jury to reasonably believe there was little to no chance that anyone wrongfully tampered with the video evidence.[122]

With minimal case law to guide authentication of video evidence, legal scholars are also recommending that electronic discovery professionals and expert witnesses must become increasingly knowledgeable of video verification to ensure proper data

---

[119] *See Chapman*, 804 F.3d at 900 (describing the backgrounds of the expert witnesses). One of the expert witnesses, Dew, had over 10 years' experience with video evidence in all forms and owned a video production company. *Id.* The second expert witness, Dickey, was an expert in "forensic evaluation and/or authentication of acoustical/visual media, including the analysis of elemental acoustics and video images contained therein." *Id.*

[120] *See supra*, History, Section A, The Pictorial Evidence Theory.

[121] *See* Mraunac, *supra* note 11 (quoting language from *Chapman* where an expert witness verified the authenticity of video surveillance evidence).

> [The expert] noted that the data appeared "consistent with surveillance recordings commonly
> associated with federal law enforcement agencies." [The expert's] procedures included verifying the frame rate of the visual recording, examining the quality of the imaging, and examining the audio embedded from the Hawk recording device. In addition, "[an] overview of HBI/VBI, color scheme, vector/waveform, embedded data, transitions and other parameters were also performed." [The expert's] report was as follows: "Data integrity checks verified the files as individually and collectively continuous. Data creation and download time/date information was also verified. All creation time and dates are sequentially uniform." Therefore, [the expert] concluded: "[the November 2, 2010, recording] does not contain any anomaly which would question its authenticity as a continuous and reliable record of the events existing therein."

Id.

[122] *See id.* (inferring that the expert's description in *Chapman* is an appropriate way to demonstrate video evidence is authentic).

collection and video analysis.[123] This includes becoming familiar with certain technical terms used in electronic discovery, such as metadata.[124] It is also vital for lawyers to educate themselves on how to confront potential deepfake evidence in the early stages of court proceedings in order to prevent the evidence from being admitted.[125] Experts recommend that attorneys take steps to verify the video's chain of custody, whether similar videos exist, and whether there are witnesses who can account for what actually happened if video tampering is suspected.[126]

It is also crucial for lawyers to consider the change in public perception as deepfake knowledge continues to grow, and how this may impact jurors' trust in video evidence.[127] Currently, there is strong favorability towards admitting video evidence, with studies demonstrating that jurors who hear oral testimony along with video testimony are 650% more likely to retain information.[128] However,

[123] *See id.* (emphasizing that it will be essential for e-discovery professionals and expert witnesses to understand deepfakes). There is the potential for additional authentication requirements to develop as deepfakes become more complex as deepfake creators use more advanced technology, and clients demand law firms to become knowledgeable in this area. *Id.*

[124] *See* Kathryn S. Lehman et al., *5 Ways To Confront Potential Deepfake Evidence in Court*, LAW360 (July 26, 2019), *archived at* https://perma.cc/GNY8-95BU (asserting various areas in discovery requests lawyers should be proficient in). *See generally* Electronic Discovery Reference Model, *EDRM Model*, EDRM (2020) [hereinafter *EDRM Model*], *archived at* https://perma.cc/FHF5-4HUH (demonstrating the various stages in the e-discovery lifecycle and defining important terms).

[125] *See* Lehman et al., *supra* note 124 (indicating that early intervention before evidence reaches the jury is key). If video evidence is admitted, it can be difficult to attack the evidence's credibility in court without getting negative reactions from the jury. *Id.*

[126] *See id.* (suggesting steps attorneys should take if they suspect that video evidence is tampered with).

[127] *See* Mraunac, *supra* note 11 (inferring various challenges that may arise when jurors understand ways in which video evidence can be tampered with).

[128] *See* Karen Martin Campbell, *Roll Tape—Admissibility of Videotape Evidence in the Courtroom*, 26 UNIV. MEMPHIS L. REV. 1445, 1447 (1996) (providing statistics on how jurors retain videotaped information at trial). Jurors who received visual testimony were 100% more likely to retain information than jurors who received only oral testimony. *Id. See also* Parry, *supra* note 27, at 185 (citing statistics on the impact of visual evidence on jurors). "Jurors often are bored, confused, and frustrated when attorneys or witnesses try to explain technical or complex material"

considering the recent increase of deepfake videos and the growing concern of deepfake influence on the 2020 presidential election and beyond, public trust in video footage may start to decline.[129] This in turn may lead to plummeting juror confidence in video evidence absent a sponsoring witness, even if it is authenticated by a judge prior to it reaching the jury.[130]

IV. Analysis

This section provides three scenarios and recommends how to confront deepfake video evidence in light of the issues that these scenarios pose.[131]

---

and having visual aids can help them retain information much better. *Id.* at 184. Jurors can retain up to 85% of information visually and in contrast only retain about 10% of what they hear. *Id.* at 185.

[129] *See* Hall, *supra* note 1, at 58 (warning that "[t]hese manufactured videos have the potential to create doubts about every recently released film."); Huston & Bahm, *supra* note 6 (stating that "the mere idea that [deepfakes] could be used to manipulate public opinion is already causing some to start questioning the validity of real events"); Thomas, *supra* note 57 (noting that deepfake technology will inevitably become "more widely commodified and accessible"); *Deepfakes in 2020 elections*, *supra* note 13 (cautioning how deepfakes may influence voters in the 2020 election); *Demystifying Deepfakes*, *supra* note 72 (emphasizing that growing concern about deepfake influence in elections draws researchers to develop applications against deepfakes); *Deepfake Detector Wins PennApps XX*, UNIV. PA. (Sept. 10, 2019) [hereinafter *PennApps XX*], *archived at* https://perma.cc/4BNT-8NEE (celebrating the first place winner at hackathon for building a deepfake detector). The University of Pennsylvania students who built the deepfake detector application did so to guard against deepfake influence on public perception in elections and to ensure that videos the public views reflect reality. *PennApps X*X, *supra*.

[130] *See* Westerlund, *supra* note 68, at 42–43 (describing how the public may begin to distrust authorities deemed reliable in the past because of deepfakes); Mirra, *supra* note 45, at 3 (cautioning that courts must be prepared for how "new technology may threaten existing and well established forms of evidence."); Hall, *supra* note 1, at 58 (contending that video may lose its value because "[t]he same accountability video that brings action can now be abused in a number of ways."); Harwell, *supra* note 13 (warning how the public may begin to generally distrust video footage because "[i]t's too much effort to figure out what's real and what's not").

[131] *See infra*, IV Sections A, B, & C (illustrating three different scenarios in which deepfake evidence will impact courtroom procedures).

A. Continuing the Silent Witness Theory if Deepfake Regulation Legislation is Passed

If Congress enacts federal legislation requiring states to enforce specific methods of video verification, courts may be able to continue using the silent witness theory without a dangerously high risk of admitting unverified false videos into evidence.[132] However, courts should only allow this application of the silent witness theory if they can determine with certainty that video manipulation occurred and that jurors will recognize indicators of tampering such as watermarks.[133] Courts continuing to apply the silent witness theory

[132] *See* H.R. 3230 § 1041 (a)–(e) (listing ways that federal legislation would require deepfake video disclosures). The federal legislation would require a watermark on any portion of the video that was digitally altered, as well as include some form of disclosure that the video is altered. *Id.* The audiovisual disclosure would require that clear text be displayed on the bottom of the video or some verbal statement detailing that the video is altered; the visual disclosure would require some clear text description on what was altered; and the audio disclosure would have at minimum one clear audio statement on what portions are altered in the video. H.R. 3230 § 1041 (c)–(e). *See also* Madison, *supra* note 17, at 711 (inferring that courts may be reluctant to abandon the silent witness theory altogether). Courts tend to abstain from barring the admission of video or photographic evidence because there is no sponsoring witness, which is why the silent witness theory was originally adopted. *Id.*

[133] *See* Lipkowitz, *supra* note 106, at 31 (cautioning that the reforms outlined in the federal deepfake legislation may not be enough to safeguard victims of deepfakes). "Watermarks are easily removable, and it is extremely difficult to track down the creators of harmful false content." *Id.* *See* Madison, *supra* note 17, at 709 (acknowledging that courts are reluctant to admit tampered video evidence). Courts will only admit video evidence that has been altered "when the jury can understand the changes in appearance that occurred between the relevant time and the time the photograph was taken." *Id.* *See* Farrell, *supra* note 14, at 2 (articulating why the silent witness theory can be applied in certain circumstances). It is only appropriate to apply the silent witness theory when the photographic or video evidence can be verified independent of a sponsoring witness and corroborated by other pieces of evidence. *Id.* *See also* Kalinski, *supra* note 11, at 815–16 (describing basic ways in which video evidence can be misleading). Even beyond deepfakes, video can easily be manipulated to make circumstances appear different than they actually are, such as angling a camera to make it seem like a head wound is worse than what the victim actually endured. *Id.* *See also* Harwell, *supra* note 13 (criticizing the current lack of research in identifying deepfakes); Venkataramakrishnan, *supra* note 56 (inferring that deepfakes are becoming more complex and will become more difficult to identify with the naked eye or ear). With little money to be made in researching how

should use more stringent requirements like the seven-prong standard because of cases' varying facts and types of visual evidence.[134] This standard provides the most rigorous verification process and offers more cohesive court guidance on the implementation of the silent witness theory.[135] If deepfake evidence frequently comes before a court, it may become necessary to amend silent witness theory protocol to require an additional qualifying witness with knowledge on deepfakes—rather than just one knowledgeable witness on the chain of custody of that piece of evidence—in order to verify that the video is an accurate representation of events to the judge.[136] It will also be

to identify deepfakes and considering the serious threat that they pose, much of the research is being funded by a government program run through the Pentagon. Harwell, *supra*.

[134] *See* Dorfman, *supra* note 11, at 20 (arguing that courts must consider adding new standards in court to ensure audio and video recordings are authentic because of deepfakes); *Videotape Evidence*, *supra* note 25, § 62 (conceding that there may be a "revival of certain foundational requirements" like the seven-prong standard because of the growing ability of the general public to edit videos). *See also* Lehman et al., *supra* note 124 (outlining an example of an interrogatory to verify the chain of custody that a lawyer may send to the opposing party).

> For the video previously produced, please provide (1) the time, place, and date of the recording was made; (2) the name and address of any individual depicted in or present at the time of the recording; (3) the name and address of any individual under whose direction and upon whose behold the recording was created; (4) the name and address of any other individual involved with the creation of the recording; (5) the steps undertaken by the in deified individuals to create the recording; and (6) the name and address of any individual who has had possession or control of the recording (either the original or a copy) since it was created.

Id.

[135] *See* Farrell, *supra* note 14 (listing the wide variety of jurisdictional practices in applying the silent witness theory); Madison, *supra* note 17, at 713 (explaining how complications can arise in verifying the authenticity of evidence when the origins of photographs cannot be verified). Madison suggests that courts should especially apply a more rigorous standard when photos of videos are taken on less sophisticated equipment such as a personal camera. Madison, *supra*.

[136] *See* Harwell, *supra* note 13 (quoting a leading deepfake researcher stating that "people do need to understand that video may not be an accurate representation of what happened"); Williams et al., *supra* note 22 (inferring the importance of a qualifying witness). Even with untampered video evidence there can be multiple interpretations, which demonstrates the importance of having a qualifying witness testify at trial. Williams et al., *supra*. *But see* Madison, *supra* note 17, at 714–15 (arguing that witnesses have imperfect memories). Even under the pictorial evidence

essential to implement training for judges to ensure that they understand the implications of technology and indicators of evidence tampering.[137] Courts will be able to measure the threat of deepfake evidence by considering the number of states adopting deepfake legislation, identifying major companies that are enacting formal policies on deepfakes, and tracking the number of individuals who were subjected to criminal penalties outlined in deepfake statutes.[138] However, given the complexities of deepfake technology and the potential for jurors to become inherently skeptical of photographic and video evidence in general, it is preferable for courts to abandon the silent witness theory and instead adopt the pictorial evidence theory.[139]

---

theory, the verifying witness may not have a completely accurate representation of events if significant time has elapsed between the event and trial. *Id.*

[137] *See* Whitney, *supra* note 42 (declaring a need to train judges about technology to ensure they properly admit evidence); *Tech Competence for Judges*, *supra* note 42 (providing statistics that two-thirds of judges say they need more training on e-discovery); Dixon, *supra* note 42 (illustrating a scenario where the authenticity of cell phone video evidence could be questioned by parties in litigation).

> (1) [A] party offers an exhibit of a cell phone video disclosed during discovery that supports the offering party's position of an agreement reached by the two parties, (2) the offering party will testify affirmatively concerning the authenticity and accuracy of the video, and (3) the opposing party will testify that he never said the words portrayed in the video.

Dixon, *supra*.

[138] *See* H.R. 3230 § 1041 (f)(1)–(f)(2) (naming the criminal and civil penalties for violating the deepfake law). Criminal penalties to disclose tampering with a video could result in 5 years imprisonment, a fine, or both. *Id.* § 1041 (f)(1). Civil penalties include a fine up to $150,000. *Id.* § 1041 (f)(2). *See Texas Deepfake Law*, *supra* note 111 (inferring that more states will create legislation banning or criminalizing deepfakes). The Texas law also provides an avenue for private individuals to seek injunctive relief if they were the victims of a deepfake video, which will infiltrate the courts with more deepfake legal controversies. *Id. See* Ingram & Ward, *supra* note 79 (inferring that more social media companies may create regulations surrounding deepfakes). Two senators wrote a letter to social media companies emphasizing that they must take a strong policy stance against deepfake videos. *Id.*

[139] *See* Dorfman, *supra* note 11, at 21 (providing potential situations of juries becoming skeptical of police body cameras and confessions in criminal cases because of deepfakes); Huston & Bahm, *supra* note 6 (warning that deepfakes are already causing people to question "the validity of real events and un-doctored video."); Amer, *supra* note 13 (cautioning that "[w]e're at a moment when we can literally no longer believe our own eyes — seeing is not necessarily believing."); LaFrance, *supra* note 79 (quoting an expert who warns that "people will not believe

B. Attacking the Silent Witness Theory if No Deepfake Regulation Legislation is Passed

If no deepfake regulation legislation is passed and courts continue to apply the silent witness theory, lawyers will need to become zealous advocates and ensure they convince a judge that there is no fathomable way video evidence should be admitted absent a qualifying witness testifying.[140] In order to accomplish this, lawyers will need to gain a complex understanding of how images can be potentially altered through deepfake technology.[141] If a lawyer

videos, just like how we do not believe photos once we're aware that tools like Photoshop exist"); Harwell, *supra* note 13 (emphasizing that deepfakes may lead to public denial of legitimate video). For example, Donald Trump said in 2016 that a leaked video of him boasting about assaulting women was doctored, leading many members of the public to falsely believe that was true. Harwell, *supra*. *See* Farrell, *supra* note 14 (defining the pictorial evidence theory). The pictorial evidence theory requires that a witness confirm the evidence is a "fair and accurate portrayal" of what they saw. *Id. See* Madison, *supra* note 17, at 707 (explaining that proponents of evidence authenticate it "by showing that the evidence accurately represents its subject."). Authentication can be easily verified by a sponsoring witness who is familiar with the subject matter of the evidence. *Id.* at 707–08.

[140] *See* Lehman et al., *supra* note 124 (describing how lawyers must attack the chain of custody of opposing counsel's video evidence if they suspect that it is a deepfake). Lawyers will need to demonstrate that there is evidence of tampering in any way possible, including identifying and locating possible witnesses who could add helpful information about the chain of custody of a video. *Id.* An example of this is locating a witness who notes issues with the metadata or a recording. *Id. See* Mraunac, *supra* note 11 (highlighting the importance of a qualifying witness testifying that the image is an accurate representation of what occurred). A Reuters reporter edited smoke after an airstrike in Lebanon to make the damage appear worse than it actually was. *Id.* The airstrike editing demonstrates the importance of having a qualifying witness to verify the true extent of events. *Id.* Even a deepfake researcher at the University of Wyoming has commented on the fact that at times, he distrusts videos that are real, and believes that videos people send him are fake because the video quality is so high. *Id. See* Williams et al., *supra* note 22 (demonstrating the importance of a qualifying witness with video evidence). Individuals would have a misunderstanding of what occurred without the journalists posting two different video angles showing that the police officers were actually dancing and not fighting. *Id. See also* Thomas, *supra* note 57 (reporting how in 2018 Sao Paulo's married governor asserted that a video of him having an orgy was a deepfake to diminish its value). The allegations could neither be confirmed nor denied. *Id.*

[141] *See* Dorfman, *supra* note 11, at 23 (asserting that detection technology cannot be the only solution to confronting deepfakes in court, and that lawyers must understand

representing a client is unable to undertake this on their own, they will need to hire an expert witness to testify as to why the evidence should not be admitted.[142] The downside to this approach will be higher costs to the client because expert witnesses can be extremely expensive and retaining them requires more attorney time that the client is billed for.[143] However, this practice may become necessary in cases that hinge on questionable video evidence and its admission to the jury.[144] Further, lawyers cannot assume that all judges fully understand the technology behind deepfake imagery, so they must use all resources

how audio and video recordings can be altered); Spivak, *supra* note 50, at 351 (inferring how complex understanding deepfake technology can be and describing technologies that run parallel to deepfake video technology); Lehman et al., *supra* note 124 (describing how video evidence is admitted under the silent witness theory after a "strong showing of authenticity and competency, including proof that the evidence was not altered.").

> [A] skilled litigator should rest assured that there will be some opportunity to attack a deepfake, even if presented pursuant to the silent witness theory. A knowledgeable witness who has identified an issue with, for example, the metadata of a recording could prove especially fruitful here, helping you to convince the judge that the other side needs to call someone to lay the chain of custody.

Lehman et al., *supra*. *See also EDRM Model*, *supra* note 112 (defining various e-discovery terms that lawyers will need to become familiar with in order to challenge potentially inauthentic video and photographic evidence).

[142] *See* Mraunac, *supra* note 11 (asserting that "expert testimony becomes relevant" when the authenticity of video evidence is questionable); Lehman et al., *supra* note 124 (articulating that a lawyer has a duty to seek expert opinion if they "have reason to suspect the video was fabricated but lack the technical knowledge to reach a conclusion"); *What Deepfakes Are*, *supra* note 49 (inferring the difficulty in staying abreast of deepfake technology if one is not a deepfake-specific researcher).

[143] *See* Dorfman, *supra* note 11, at 20 (warning of the "steep costs" of litigation that may occur because of deepfakes); Lehman et al., *supra* note 124 (inferring that experts are expensive to hire and the cost should be carefully evaluated in light of the value of the case).

[144] *See* Scott v. Harris, 550 U.S. 372, 378–81 (2007) (illustrating the huge impact one piece of video evidence can have in the outcome of a case); Parry, *supra* note 27, at 185 (warning how "there is absolutely no way the average juror could tell the difference between a doctored and a pure photo."); Granot et al., *supra* note 12 at 94 (inferring the immense power visual evidence has in court cases through jurors evaluating video with "naïve realism"). *See also* Lehman et al., *supra* note 124 (providing advice on what lawyers should do if questionable video evidence is admitted). Lawyers should aggressively attack the chain of custody through cross-examination if questionable video evidence is admitted. *Id.* Once jurors realize the authenticity of the video is in question, they will consider whether or not they should assign serious weight to it in deliberations. *Id.*

available to communicate the technology's potentially grave implications on a case.[145] The lawyer will need to question the authenticity of the opposing party's video evidence and convince the judge that, at a bare minimum, the opposing party must present a witness at trial for cross-examination about the process used to collect the admitted evidence.[146] Some may argue that this is too excessive of a process for seemingly valid evidence, but given the expansion of deepfake videos into originally unimagined areas, it is essential for lawyers to be vigilant for their clients when encountering visual evidence.[147]

---

[145] *See Tech Competence for Judges*, *supra* note 42 (quoting statistics demonstrating that two-thirds of judges feel they need more education on e-discovery); Harris, *supra* note 7, at 110 (noting how "[d]espite the transformative nature" of deepfakes judges are relatively inexperienced in dealing with this type of technology); Pfefferkorn, *supra* note 42 (cautioning that deepfakes could infiltrate trial courts). "Points where deepfakes could infect a court case run the gamut from clients who fabricate evidence in order to win, to fake videos ending up in archives that have historically been considered trustworthy." Pfefferkorn, *supra*.

[146] *See* Lehman et al., *supra* note 124 (recommending that litigators should request an authentication witness from the opposing party that may be cross-examined). The main idea behind the cross-examination is to call the jury's attention to the fact that there is a disagreement over the evidence's authenticity. *Id.* *See also* Delfino, *supra* note 7, at 895 (stating that "[t]o protect victims of deepfakes and to prevent the negative societal consequences they cause, the laws need to keep pace with this technology."); Huston & Bahm, *supra* note 6 (listing ways in which deepfake technology can be abused to manipulate public perception); Harwell, *supra* note 13 (indicating that researchers "remain vastly overwhelmed" as deepfake technology advances); Goggin, *supra* note 8 (declaring the widespread use of deepfakes and that researchers cannot keep up with new technologies used to create them); Stankiewicz, *supra* note 10 (emphasizing how difficult it will become to identify deepfake videos in just a few months).

[147] *See* Chesney & Citron, *supra* note 9, at 1763 (cautioning how accessible photo and video manipulation technology is to the general public); Harris, *supra* note 7, at 123–24 (warning about the dangers of hyper-realistic deepfakes and the need for the legal system to take action); Stupp, *supra* note 116 (providing an alarming case where a CEO was fooled into wiring millions of dollars to a scammer using deepfake voice recordings); *USPTO*, *supra* note 7 (articulating the growing problem of fake patent and trademark applications); Jee, *supra* note 48 (reporting on the use of deepfake technology by an Indian political party to influence voters in an election). "Deepfakes will become so life-like that they will be indistinguishable from actual videos. This is an inevitable consequence of artificial intelligence and machine learning technologies. The law should be equipped to handle this impending problem." Harris, *supra*. *See also* Ingram & Ward, *supra* note 79 (commenting on the scale of the deepfake analysis problem by noting that even Facebook is outsourcing research to combat fake videos on its site).

C. Adopting the Pictorial Evidence Theory if No Deepfake Regulation Legislation is Passed

Absent significant deepfake legislation, courts should adopt the pictorial evidence theory to combat heightened public skepticism of photographic and video evidence.[148] There have already been instances of high-profile criminal activity involving the use of deepfakes, such as the hackers who used the technology to impersonate a parent company executive and convince one of its subsidiaries to transfer nearly $250,000 into a bank account controlled by the perpetrators.[149] The Chief Executive Officer who transferred the funds was under the genuine belief that he was speaking with the proper executive and participating in an authorized transaction.[150] This example demonstrates just one of the potential threats posed by deepfakes to public trust in audio and video recordings.[151] Because growing media coverage about the dangers of deepfakes will likely lead the public—which includes jurors—to view videos with skepticism, courts should consider adopting the pictorial evidence

[148] *See* Harwell, *supra* note 13 (stressing that hysteria and fake videos will change how the public considers videos and photos); Goggin, *supra* note 8 (noting how the deepfake movement has grown substantially since its initial rise in 2017); Stankiewicz, *supra* note 10 (emphasizing that everyday people will have access to "perfectly-real" deepfake videos in six months to one year). *See also* Delfino, *supra* note 7, at 890 (highlighting the alarming ease at which deepfake content can be created). "Deepfake technology has evolved so quickly that an app designed to create deepfakes is now widely available. The app lowers the technical threshold required to create such images and videos, which will likely make . . . [them] . . . more prevalent." *Id. See also USPTO*, *supra* note 7 (providing details about a new reporter program created by the United States Patent and Trademark Office). Artificially generated images are becoming so frequent with patent applications that the USPTO created a new program for the public to report any images they identify as fake. *Id.*

[149] *See* Stupp, *supra* note 116 (describing deepfake criminal activity that led to money being unknowingly transferred to a criminal).

[150] *See id.* (emphasizing that the employee transferred the money in good faith). The audio in this case was so convincing that it fooled even someone who worked closely with the person that the deepfake was imitating. *Id.*

[151] *See* Mirra, *supra* note 45, at 16 (warning about the impact voice alteration software can have on verdicts in criminal cases). "The entire course of a defendant's life could be altered with some quick changes made on [voice editing software] to a voice recording that provides a pivotal piece of evidence at trial." *Id. See also Eckert Seamans*, *supra* note 45 (detailing a scam where a lawyer father almost wired $9,000 to a scammer believing that it was his son).

standard for all video evidence to ensure that jurors properly consider the weight of evidence.[152] The pictorial evidence theory requires a witness to testify before the jury that the photographic or video evidence is a fair and accurate representation of what occurred.[153] Having the witness on the stand to verify that the visual evidence is accurate will allow an initially skeptical jury to view the evidence as a true image and not question whether it was artificially generated.[154]

At a minimum, courts should always require a witness to authenticate the process under which the video was obtained before

[152] *See* Westerlund, *supra* note 68, at 42–43 (cautioning that the public will become skeptical of normally trustworthy authorities because of deepfakes); Amer, *supra* note 13 (noting the role that the media plays in directing public attention to deepfakes); LaFrance, *supra* note 79 (alleging that deepfakes are dangerous because "[p]eople are already fooled by doctored photos, impostor accounts on social media, and other sorts of digital mimicry all the time"); Engler, *supra* note 10 (asserting that deepfakes are dangerous to public trust in digital content). "Deepfakes pose a significant problem for public knowledge. Their development is not a watershed moment—altered images, audio, and video have pervaded the internet for a long time—but they will significantly contribute to the continued erosion of faith in digital content." Engler, *supra*. *See also* FED. R. EVID. 403 (providing circumstances under which a judge may decide to exclude relevant evidence). "The court may exclude relevant evidence if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence." *Id.* *See also* Tanford, *supra* note 14 (illustrating reasons why evidence has probative value and reasons probative value can be substantially outweighed by unfair prejudice).

[153] *See* Farrell, *supra* note 14, at 2 (explaining how witness testimony is used to authenticate events illustrated in photographic and video evidence under the pictorial evidence theory).

[154] *See* U.S. v. Chapman, 804 F.3d 895, 900–01 (7th Cir. 2015) (providing an example of how courts are already willing to allow multiple experts if there is a showing that evidence was tampered with). Chapman introduced two experts to testify as to the authenticity of the video expert and wished to call a third. *Id.* at 900. The lower court denied the motion, but "suggested that it might have considered the appointment if Chapman had indicated any flaw in the software or system the government used to transfer the recording from the Hawk recording device onto the DVD." *Id.* at 901. *See* Engler, *supra* note 10 (warning of a "threefold" effect on public perception of information because of deepfakes). As a result of deepfake technology, the public will: (1) have a "visceral reaction" to fake information and spread it more quickly; (2) have more difficulty in ascertaining what stories are true; and (3) accuse true sources of being fake because of disagreeable content. *Id.* *See also USPTO*, *supra* note 7 (describing how the USPTO is becoming skeptical of trademark applications). The USPTO is relying on the public to draw attention to patent or trademark applications that it believes are artificially generated because there has been such an influx of fake applications. *Id.*

the jury, even in typically routine circumstances such as surveillance videos.[155] Critics may argue that always applying the pictorial evidence theory is an unnecessarily high standard because most deepfakes are pornographic.[156] However, with an over 80% increase in deepfakes in just one year and recent uses of deepfakes to influence business and politics, it is dangerous to assume deepfakes will not become a legal problem, especially in high-profile cases.[157] People are so influenced by video evidence that they can be convinced that they committed an act that they did not do.[158] One study conducted by a bank had participants view doctored videos of themselves stealing money.[159] After watching the doctored video, the participants would confess to taking money, despite knowing that they had not stolen

[155] *See* Parry, *supra* note 27, at 186 (calling upon courts to issue stricter evidence admission standards because of the ease at which digital evidence can be altered); Spivak, *supra* note 50, at 342 (emphasizing that Photoshop technology makes the public skeptical of visual evidence). *See also Chapman*, 804 F.3d at 900 (using helpful language in considering the value of expert witness testimony in authenticating the process by which evidence is collected).

[156] *See* Metz, *supra* note 50 (stating that 96% of the deepfakes that Deeptrace found online were pornography). *But see* Attestiv, Inc., *supra* note 80 (announcing how the company is launching an application to detect fake media "[i]n a world rampant with digital fraud and rapidly emerging deepfake technology.").

[157] *See* Hall, *supra* note 1, at 75 (forewarning that "[d]eepfake videos will only become more of a problem."); Henderson, *supra* note 44, at 1148–49 (contending that courts already confront doctored photographic evidence in the context of child pornography cases); Huston & Bahm, *supra* note 6 (emphasizing that confronting deepfakes will require "a whole of society approach" where multiple organizations and entities proactively consider how to handle the technology); Stupp, *supra* note 116 (discussing a case where a CEO wired millions to a scammer using deepfake voice recordings to pose as his supervisor); LaFrance, *supra* note 79 (cautioning on the danger of deepfakes influencing the public by stating "[i]magine the confusion that might surround a convincing video of the president . . . say[ing] something he never actually said."); Goggin, *supra* note 8 (listing incidents of deepfake videos targeting politicians Nancy Pelosi, Barack Obama, and Alexandria Ocasio-Cortez); Metz, *supra* note 50 (providing statistics from the Deeptrace study stating that there was an 84% increase in deepfake videos since December 2018); Fagan, *supra* note 56 (reporting on how a deepfake video of Barack Obama calling Donald Trump a "dipshit" quickly reached millions of people); Nelson & Simek, *supra* note 56 (warning about the easy access the public has to deepfake creation technology through apps like DeepNude).

[158] *See* Granot et al., *supra* note 12, at 97–98 (emphasizing the impact video evidence had in convincing participants they committed a crime in a study conducted by a bank).

[159] *See id.* (illustrating the background of the bank study).

anything.[160] If deepfake videos can change peoples' minds about what they did themselves, it only becomes more dangerous when used by jurors to evaluate situations that they were not a part of.[161] Due to First Amendment concerns, it is unlikely that deepfakes will ever be banned entirely; therefore, it is of the utmost importance that courts adapt to this technological threat before it infiltrates the legal system and creates widespread juror doubt in video evidence.[162]

V. Conclusion

Misuse of deepfake technology is an imminent threat to society, and there is no doubt it will eventually impact the courtroom. The number of researchers creating tools to identify doctored videos is severely outpaced by the number of people creating them. Deepfake videos are projected to quickly become so convincing that they will be unidentifiable with the naked eye. It is of the utmost importance that lawyers and the legal system as a whole consider how to handle this threat before it undermines the justice system. While Congress is attempting to take some action, the legal system should not put faith in legislation alone. More jurors will become aware of deepfakes and without proactive judicial safeguards in place, they will resultantly be more skeptical of visual evidence in court proceedings. Allowing the continuous inconsistent application of the silent witness theory across jurisdictions will only fuel the growing mistrust of visual evidence. Courts should adopt the pictorial evidence theory and abolish use of

[160] *See id.* (stating the results of the bank video study).

[161] *See id.* (inferring how dangerous visual evidence can be in changing peoples' perceptions of events).

[162] *See* Harris, *supra* note 7, at 128 (emphasizing the need for the legal system to be proactive in its dealings with deepfakes because "technology is only going to improve"); Smith-Roberts, *supra* note 5, at 123 (articulating how difficult it is to restrict "even false speech" in the United States because of First Amendment protections); Hall, *supra* note 1, at 62 (explaining how "[m]any deepfake videos would be protected by the First Amendment as free expression under the defense of parody or satire . . . ."); Spivak, *supra* note 50, at 356–64 (discussing First Amendment concerns that deepfake legislation raises). First Amendment challenges to deepfakes will likely prevail, unless the deepfake falls under the category of obscenity or child pornography. Spivak, *supra*. *See* Feiner, *supra* note 108 (contrasting the tension between free speech and the threat of deepfakes). Facebook stated the following on fake videos being posted on its site: "'[t]here's a tension here: we work hard to find the right balance between encouraging free expression and promoting a safe and authentic community . . . . '" *Id.*

the silent witness theory in order to ensure that admitted video evidence is authentic and that jurors place appropriate weight in video evidence. Deepfakes threaten the foundation that our justice system is built upon—that evidence leads to the truth and a just outcome.