ARTICLE:

"DEEPFAKES" IN THE COURTROOM

RIANA PFEFFERKORN[*]

ABSTRACT

Seeing is believing—but for how long? At present, people attach a lot of probative weight to images and videos. They're taken at face value as evidence that an event occurred as alleged. The advent of so-called "deepfake" videos might change that. Thanks to advances in artificial intelligence, it is now possible to create a genuine-looking video that makes real people appear to do and say things they never did or said. Software for creating deepfake images, video, and audio is already freely available online and fairly easy to use. As the technology rapidly advances, it will become harder for humans and computers alike to tell a fake video from a real one.

Inevitably, deepfakes will start coming up in the courtroom context. This Article surveys the ramifications of deepfakes for pre-trial and trial practice, including authentication of evidence, professional responsibility, and a potential

[*] Associate Director of Surveillance and Cybersecurity, Center for Internet and Society, Stanford Law School. A version of this article originally appeared in the September 2019 issue of *Washington State Bar News* (formerly *NWLawyer*), published by the Washington State Bar Association (WSBA), and is reprinted with the permission of the WSBA. Thanks to the organizers of this Symposium for inviting me to contribute this article; to Microsoft for generously providing funding support for it; to Al Gidari, Stevie DeGroff '19, and George Brown '20 at Stanford Law School, for thoughtful discussions that helped inform this article; and to Tait Graves for providing feedback on a draft version. All errors are, of course, my own. This article is dedicated to the memory of Jim Tyre.

"reverse CSI effect" on juries primed to question even authentic evidence in an era of disinformation and "fake news." Fortunately, courts are no stranger to the phenomenon of evidence tampering and forgery. The rules of evidence have long imposed authentication requirements to help screen out fakes. I argue that those requirements are sufficient as-is to deal with deepfakes, and that raising the bar for authenticating video evidence would do more harm than good. Although it may prove costly, courts will be able to handle the challenges posed by deepfakes as they have ably handled previous generations of inauthentic evidence.

> Photographs furnish evidence. Something we hear about, but doubt, seems proven when we're shown a photograph of it. In one version of its utility, the camera record incriminates . . . . In another version of its utility, the camera record justifies. A photograph passes for incontrovertible proof that a given thing happened. The picture may distort; but there is always a presumption that something exists, or did exist, which is like what's in the picture.[1]

INTRODUCTION

"Deepfake" videos, images, and audio stand poised to add a new chapter to trial courts' history of dealing with inauthentic evidence. A portmanteau of "deep learning" and "fake,"[2] so-called "deepfake" software programs use artificial intelligence (AI) to produce forged videos of real (and even entirely fabricated[3]) people that appear genuine, making them appear to do and say things they never did or said.[4] The more video and audio footage of real people that can be fed into the system, the more convincing the result. Software for creating deepfakes is already freely available online and fairly easy for anyone to use. As the software's usability and the videos' verisimilitude keep improving over time, it will become harder for laypeople, as well as computer systems, to tell real from fake.

The advent of deepfakes will affect American courts in multiple ways. To date, the legal scholarship around deepfakes has focused on potential causes of action for those harmed by them, possible remedies, and the propriety of additional regulation.[5] But deepfakes will not only provide the grounds for filing

[1] SUSAN SONTAG, "In Plato's Cave," ON PHOTOGRAPHY 3 (RosettaBooks LLC 2005) (1977).

[2] Noelle Martin with Daniella Scott, *Deepfake Porn Nearly Ruined My Life*, ELLE (Feb. 6, 2020), https://www.elle.com/uk/life-and-culture/a30748079/deepfake-porn/. Deep learning is a subfield of machine learning, which in turn is a subfield of artificial intelligence. DEEP LEARNING, https://en.wikipedia.org/wiki/Deep_learning (last visited Mar. 13, 2020); MACHINE LEARNING, https://en.wikipedia.org/wiki/Machine_learning (last visited Mar. 13, 2020).

[3] James Vincent, *ThisPersonDoesNotExist.com uses AI to Generate Endless Fake Faces*, THE VERGE (Feb. 15, 2019), https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan.

[4] *See* Charles Q. Choi, *AI Creates Fake Obama*, IEEE SPECTRUM (July 12, 2017), https://spectrum.ieee.org/tech-talk/robotics/artificial-intelligence/ai-creates-fake-obama. Software programs for creating audio-only deepfakes are also available. *See* Kyle Wiggers, *Resemble AI Launches Voice Synthesis Platform and Deepfake Detection Tool*, VENTUREBEAT (Dec. 17, 2019), https://venturebeat.com/2019/12/17/resemble-ai-launches-voice-synthesis-platform-and-deepfake-detection-tool/ (describing programs including Lyrebird, Deep Voice, and Resemble, and noting that "only a few minutes — and in the case of state-of-the-art models, a few seconds — are required to imitate a subject's prosody and intonation with precision").

[5] *See* Section III *infra*.

suit. They will also come up in a range of civil and criminal cases as just another piece of evidence that a party seeks to introduce.

This Article surveys the ramifications of deepfakes for pre-trial and trial practice, including authentication of evidence, professional responsibility, and the potential for a "reverse CSI effect" on jurors who, in the era of "fake news," may be primed to doubt the veracity of even legitimate evidence. I argue that imposing a heightened bar for authentication of video evidence is unwarranted. I predict that, while it may prove costly, courts will be able to handle the challenges posed by deepfakes as they have handled potentially inauthentic evidence in the past.

I. WHAT ARE DEEPFAKES?

In the summer of 2017, a team of computer scientists at the University of Washington ("UW") caused a stir by building algorithms that allowed them to generate a realistic, but phony, video of former president Barack Obama based on actual audio and video clips.[6] The following spring, comedian Jordan Peele used a software tool that was available for free online to create another ersatz video of President Obama. This one seemed to show President Obama speaking words that were in fact being spoken by Peele, as revealed in a split-screen view[7] evocative of the unveiling of "that man behind the curtain" who was manipulating the image of the Wizard of Oz.[8]

These kinds of videos are popularly known as "deepfakes."[9] The UW team created its "deepfake" Obama video by training a neural network[10] to "translate different audio sounds into basic mouth shapes," after which the team could "realistically superimpose and blend those mouth shapes and textures on an existing reference video of" Obama.[11] The more video and audio footage of real

[6] Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1760 (2019) (footnotes omitted); Jennifer Langston, *Lip-Syncing Obama: New Tools Turn Audio Clips into Realistic Video*, UW NEWS (July 11, 2017), https://www.washington.edu/news/2017/07/11/lip-syncing-obama-new-tools-turn-audio-clips-into-realistic-video/.

[7] Craig Silverman, *How to Spot a Deepfake Like the Barack Obama-Jordan Peele Video*, BUZZFEED NEWS (Apr. 17, 2018), https://www.buzzfeed.com/craigsilverman/obama-jordan-peele-deepfake-video-debunk-buzzfeed.

[8] THE WIZARD OF OZ (Metro-Goldwyn-Mayer 1939).

[9] *See* Chesney & Citron*, supra* note 6, at 1758, 1760.

[10] An artificial neural network is a computing system "vaguely inspired by" animal brains, which "'learn[s]' to perform tasks by considering examples, generally without being programmed with task-specific rules." ARTIFICIAL NEURAL NETWORK, https://en.wikipedia.org/wiki/Artificial_neural_networks (last visited Mar. 13, 2020).

[11] Langston, *supra* note 6.

people that can be fed into the AI system’s deep-learning algorithms, the more convincing the result.[12]

While the UW team’s approach involved a single neural network,[13] many deepfake systems are built on “generative adversarial networks,” or GANs. Pioneered in 2014, GANs “are two-part AI models consisting of a *generator* that creates samples [of video, images, or audio] and a *discriminator* that attempts to differentiate between the generated samples and real-world samples.”[14] The generator “draws on a dataset to produce a sample that mimics the dataset,” and then the discriminator “assesses the degree to which the generator succeeded.”[15] Iteratively “pitting two algorithms against each other” improves both sides’ performance: as the discriminator gets better at spotting fakes, it provides feedback to the generator, which learns from its mistakes and produces more realistic output, which in turn is then fed back into the generator, and so on.[16]

At present, the most cutting-edge deepfake technology arises from academic and corporate research settings.[17] While not as high-quality in output,[18] however, software programs for creating deepfakes (such as Peele’s[19]) have been freely available online since at least 2017,[20] and they’re fairly easy for

[12] President Obama was a particularly suitable subject for generating a high-quality deepfake, due to the exceptionally high volume of public-domain video of him available to be fed into the researchers’ system for analysis. Langston, *supra* note 6; Choi, *supra* note 4.

[13] *See* Choi, *supra* note 4.

[14] Kyle Wiggers, *Generative Adversarial Networks: What GANs Are and How They’ve Evolved*, VENTUREBEAT (Dec. 26, 2019), https://venturebeat.com/2019/12/26/gan-generative-adversarial-network-explainer-ai-machine-learning/ (crediting Ian Goodfellow as the “father” of GANs, and IBM’s Arthur Samuel, who allegedly popularized the term “machine learning,” as GANs’ “grandfather.”). Goodfellow coauthored the “seminal” paper that “describe[d] the first working implementation of a generative model based on adversarial networks.” *Id.* (citing Ian Goodfellow *et al.*, *Generative Adversarial Nets*, ARXIV (June 10, 2014), https://arxiv.org/pdf/1406.2661.pdf).

[15] Chesney & Citron, supra note 6, at 1760 (footnotes omitted).

[16] Wiggers, *supra* note 14.

[17] For example, researchers from Stanford University, the Max Planck Institute for Informatics, Princeton University, and Adobe Research created software that lets users edit a video so as to “change the words coming right out of somebody’s mouth,” but “[t]his work is just at the research stage right now and isn’t available as consumer software.” James Vincent, *AI Deepfakes Are Now as Simple as Typing Whatever You Want Your Subject to Say*, THE VERGE (June 10, 2019), https://www.theverge.com/2019/6/10/18659432/deepfake-ai-fakes-tech-edit-video-by-typing-new-words (adding “but it probably won’t be long until similar services go public”).

[18] Silverman, *supra* note 7 (“[I]t still requires a decent amount of skill, processing power, and time to create a really good ‘deepfake.’”).

[19] *Id.* (naming FakeApp as the “free tool” used).

[20] *See* Martin & Scott, *supra* note 2 (pinning 2017 as the year “when the [online discussion forum website] Reddit community began creating deepfakes for themselves”); Silverman, *supra* note 7 (describing the free FakeApp software tool used to create the Peele video).

anyone to use.[21] The technology's accessibility, usability, and the verisimilitude of its output will keep improving over time.[22] Consequently, it will become harder and harder for people—and AI systems themselves—to tell real videos from fake ones.

That means deepfake detection is a topic of significant interest. The U.S. government, academia, nonprofits, and the tech industry have all launched initiatives (sometimes in collaboration with each other) to push forward the state of technology for detecting deepfakes.[23] Not everyone, however, is sanguine about the future of detection: my Stanford colleagues Dan Boneh and Andrew Grotto have opined that "in the long-run [deepfake detection] is likely to be a losing battle or at best a stalemate."[24]

If proving which videos are fake becomes too difficult, then maybe it would be easier to establish which videos aren't—to prove an affirmative rather than a

[21] *See* Timothy B. Lee, *I Created My Own Deepfake—It Took Two Weeks and Cost $552*, ARS TECHNICA (Dec. 16, 2019), https://arstechnica.com/science/2019/12/how-i-created-a-deepfake-of-mark-zuckerberg-and-star-treks-data/ (describing author's use of DeepFaceLab and Faceswap software programs).

[22] Cade Metz, *Internet Companies Prepare to Fight the 'Deepfake' Future*, N.Y. TIMES (Nov. 24, 2019), https://www.nytimes.com/2019/11/24/technology/tech-companies-deepfakes.html ("The technology has improved at a rate that surprises A.I. experts, and there is little reason to believe it will slow.").

[23] For example, the Defense Advanced Research Projects Agency (DARPA) has instituted the Media Forensics (MediFor) program "to attempt to level the digital imagery playing field, which currently favors the manipulator"; if successful, "the MediFor platform will automatically detect manipulations," among other goals. DARPA, Media Forensics (MediFor), https://www.darpa.mil/program/media-forensics (last visited Feb. 28, 2020); Will Knight, *The Defense Department Has Produced the First Tools for Catching Deepfakes*, MIT TECH. REV. (Aug. 7, 2018), https://www.technologyreview.com/s/611726/the-defense-department-has-produced-the-first-tools-for-catching-deepfakes/. Google has released datasets of fake audio and video in order to assist the deepfake detection research efforts being carried out by others, including academics in Germany and Italy. Nick Dufour, *Contributing Data to Deepfake Detection Research*, GOOGLE AI BLOG (Sept. 24, 2019), https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html (last visited Feb. 28, 2020). Between December 2019 and March 2020, Facebook, Amazon, Microsoft, academics from eight universities, and the nonprofit Partnership on AI held the "Deepfake Detection Challenge," where "researchers around the world [vied] to create automated tools that can spot fraudulent media." Eliza Strickland, *Facebook AI Launches Its Deepfake Detection Challenge*, IEEE SPECTRUM (Dec. 11, 2019), https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/facebook-ai-launches-its-deepfake-detection-challenge (last visited Mar. 2, 2020). As a counterpoint to the free accessibility of tools for generating deepfakes, deepfake detection software has also been released publicly for anyone to access. *E.g.*, Michael Valeriani, Shallow, on GitHub, https://github.com/mvaleriani/Shallow (last visited Feb. 28, 2020).

[24] Dan Boneh, Andrew J. Grotto, *et al.*, *How Relevant is the Turing Test in the Age of Sophisbots?* at *3, ARXIV (Aug. 30, 2019), https://arxiv.org/pdf/1909.00056.pdf (endnote omitted).

negative.[25] To that end, various tools have been introduced for authenticating video recordings. The goal is to vouch for video recordings' authenticity, and to make it evident if a video has been manipulated in some way from the original.[26]

As the battle over authentic video continues, it's not clear which camp will eventually prevail. Will deepfake detection and video authentication technologies be able to keep up with the technology for generating fakes? Or are we about to enter an era where it's no longer possible for either humans or machines to spot fakes? Only one thing seems certain: we won't see a definitive end to this "cat and mouse arms race"[27] anytime soon.

II. DEEPFAKES' IMPACT ON THE LAW

Deepfakes are still a nascent topic in the law. While deepfake-related research in computer science dates back to at least mid-2014,[28] it took a few years for the technology to get to the point where it began to concern legal scholars[29] and policymakers.[30] So far, most of their efforts have focused on what we might call the question of containment: how to prevent, mitigate, and punish the abuse of

[25] *Ticks or It Didn't Happen: Confronting Key Dilemmas in Authenticity Infrastructure for Multimedia*, at 6, WITNESS (December 2019), https://lab.witness.org/ticks-or-it-didnt-happen/ (hereinafter *Ticks*) ("The idea is that if you cannot detect deepfakes, you can, instead, authenticate images, videos and audio recordings at their moment of capture.").

[26] *Id.* at 10 ("There is a growing sense of urgency around developing technical solutions and infrastructures that can provide definitive answers to whether an image, audio recording or video is 'real' or, if not, how it has been manipulated, re-purposed or edited since the moment of capture."). The *Ticks* report provides an overview of authentication tools that were in use or in development as of October 2019. *Id.* at 11–15.

[27] *Ticks*, *supra* note 25, at 47; *see also* Metz, *supra* note 22.

[28] *See* Ian Goodfellow *et al.*, *Generative Adversarial Nets*, ARXIV (June 10, 2014), https://arxiv.org/pdf/1406.2661.pdf.

[29] *See* Robert Chesney & Danielle Citron, *Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?*, LAWFARE (Feb. 21, 2018), https://www.lawfareblog.com/deep-fakes-looming-crisis-national-security-democracy-and-privacy. This piece, featured on a well-respected national-security blog, kicked off the legal scholarship on deepfakes and was subsequently expanded into Chesney and Citron's thorough and prescient law review article. *See* Chesney & Citron, *supra* note 6.

[30] *See, e.g.*, *Cybersecurity and California Elections* before a Joint Informational Hearing, Assemb. Comm. on Elections & Redistricting and S. Comm. on Election & Constitutional Amendments, 2018 Leg. [inset session here] (Cal. 2018) (testimony of Andrew Grotto), https://cisac.fsi.stanford.edu/docs/andrew-grotto-testimony-cybersecurity-and-california-elections; Kaveh Waddell, *Lawmakers Plunge into 'Deepfake' War*, AXIOS (Jan. 31, 2019), https://www.axios.com/deepfake-laws-fb5de200-1bfe-4aaf-9c93-19c0ba16d744.html (reporting that federal legislators had begun "invit[ing] legal scholars to privately brief their staff on deepfakes").

deepfake technology for harmful purposes.[31] When deepfakes cause harm—whether on a small scale (e.g., to an individual person[32]) or large scale (e.g., to national security[33])—how should the law respond?[34] What existing civil and criminal laws could be invoked to redress those harms, what remedies are available to those injured, and what new regulations may be called for?[35] If new

[31] 2019 Cal. Stat. ch. 493 § 3 (AB 730) (expanding California's Truth in Political Advertising Act, Cal. Elec. Code § 20010, to cover malicious use of deepfakes in campaign materials); 2019 Cal. Stat. ch. 491 § 1 (AB 602) (expanding California's "revenge porn" law to include deepfakes, Cal. Civ. Code § 1708.86); 2019 Va. Acts 490 (similarly amending Virginia's revenge porn law, Va. Code Ann. § 18.2-386.2). *See also* K.C. Halm, Ambika Kumar Doran, Jonathan Segal, and Caesar Kalinowski IV, *Two New California Laws Tackle Deepfake Videos in Politics and Porn*, DAVIS WRIGHT TREMAINE (Oct. 14, 2019), https://www.dwt.com/insights/2019/10/california-deepfakes-law; Kirsten Korosec, *'Deepfake' Revenge Porn Is Now Illegal in Virginia*, TECHCRUNCH (July 1, 2019), https://techcrunch.com/2019/07/01/deepfake-revenge-porn-is-now-illegal-in-virginia/.

[32] The vast majority of existing deepfake videos are non-consensual pornographic ones that insert real women's likenesses (both celebrities and private individuals) into sexual situations in which they never actually appeared. *The State of Deepfakes: Landscape, Threats, and Impact* foreword & 2, DEEPTRACE LABS (September 2019), *report available at* https://deeptracelabs.com/mapping-the-deepfake-landscape/ (report by artificial-intelligence company Deeptrace found that 96% of the nearly 15,000 deepfake videos counted were pornographic, of which 99% mapped female celebrities' faces onto actresses in pornographic videos); *see also* Martin & Scott, *supra* note 2 (account by young Australian woman of being the subject of a pornographic deepfake in 2016, then successfully campaigning to make "image-based abuse" a criminal offense in Australia the following year). The problem of nonconsensual faked pornographic photos of women is almost as old as photography itself: by the late nineteenth century, women who sat for photo portraits were already being warned that "an ungentlemanly photographer could use the negative . . . [for] base purposes . . . includ[ing] making composites that featured the sitter's head on a scantily dressed body—or worse." Lynn Berger, Photography Distinguishes Itself: Law and the Emerging Profession of Photography in the Nineteenth-Century United States, 248 (Feb. 26, 2016) (unpublished Ph.D. dissertation, Columbia University) (https://academiccommons.columbia.edu/doi/10.7916 /D8222TM3) (citation and parentheses omitted).

[33] Nat'l Def. Auth. Act for Fiscal Year 2020, Pub. L. No. 116-92 § 5709, 133 Stat. 1198 (2019) (first federal deepfake legislation, requiring Director of National Intelligence to study the problem and issue a report). The law, which was part of an omnibus defense spending bill, focuses on foreign countries' weaponization of deepfakes and on stimulating research and development efforts for deepfake detection technologies. Matthew F. Ferraro, Jason C. Chipman, and Stephen W. Preston, *First Federal Legislation on Deepfakes Signed into Law*, WILMERHALE (Dec. 23, 2019), https://www.wilmerhale.com/en/insights/client-alerts/20191223-first-federal-legislation-on-deepfakes-signed-into-law.

[34] The Maryland Law Review addressed this question, examining such areas of law as the rights of privacy and publicity, free expression, and national security. Symposium, *Truth Decay: Deep Fakes and the Implications for Privacy, National Security, and Democracy*, 78 MD. L. REV. 882–966 (2019).

[35] *See generally* Chesney & Citron, *supra* note 6 at 1788–1808 (reviewing a range of potential civil, criminal, and regulatory responses to deepfakes); *see also* Rebecca A. Delfino,

laws are indeed appropriate, what other frameworks (such as the First Amendment) might constrain their scope?[36]

In contrast to the legal scholarship exploring the substantive areas of law that deepfakes implicate, this Article focuses on a more mundane procedural matter: deepfakes' implications for evidentiary proceedings in court.[37] We may safely assume that the ready availability of deepfake tools, and antisocial uses thereof, will continue irrespective of how the law may attempt to contain, regulate, and punish them.[38] If deepfakes are here to stay, then the law must be ready to respond to their effects on our legal system.

The foreseeable effects will be both direct and indirect. Directly, deepfakes will cause some additional caseload in the courts. Those cases will be about the deepfakes themselves, such as where a deepfake video gives rise to a tort claim.[39] Indirectly, however, deepfakes will also affect the courts by playing a

Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act, 88 FORDHAM L. REV. 887 (2019).

[36] Several recent scholarly articles have addressed First Amendment issues with regulating deepfakes. *E.g.*, Marc Jonathan Blitz, *Lies, Line Drawing, and (Deep) Fake News*, 71 OKLA. L. REV. 59, 86–115 (2018); Rebecca Green, *Counterfeit Campaign Speech*, 70 HASTINGS L.J. 1445, 1476–85 (2019); Jared Schroeder, Free Expression Rationales and the Problem of Deepfakes within the E.U. and U.S. Legal Systems *11–22 (Dec. 13, 2019) (unpublished manuscript) (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3503617).

[37] This article builds on my previous, shorter writing and interviews on the same topic. Riana Pfefferkorn, *Too Good to Be True? "Deepfakes" Pose a New Challenge for Trial Courts*, 73 NW LAWYER 7, 23–25 (Sept. 2019), https://wabarnews.wsba.org/wabarnews/sept_2019/MobilePagedReplica.action?pm=2&folio=22#pg24. *See also Ticks*, *supra* note 25, at 28, 67 (interviewing me as part of a comprehensive examination of the various dilemmas posed by deepfakes and how they intersect, with a focus on defending human rights); Vanessa Blum, *Coming Soon to a Courtroom Near You? What Lawyers Should Know About Deepfake Videos*, THE RECORDER (Mar. 14, 2019), https://www.law.com/therecorder/2019/03/14/coming-soon-to-a-courtroom-near-you-heres-what-you-need-to-know-about-deepfake-videos/ (interviewing me prior to the publication of my WSBA article).

[38] The impossibility of putting the genie back in the bottle does not undermine the important work of those engaged in crafting potential responses. Deepfakes will—and already do—directly harm people, particularly women and girls. *See The State of Deepfakes*, *supra* note 32. Legal, market, and societal measures may be able to mitigate those harms, ideally without unduly impairing prosocial uses of deepfakes or other values such as free speech and privacy. *See* Chesney & Citron, *supra* note 6, at 1769–71, 1788–92, 1814–17 (noting the beneficial uses of deepfakes, as well as the freedom-of-expression and privacy concerns raised by potential responses to deepfakes).

[39] Doctored photographs have given rise to tort actions before. For example, in *Morsette v. "The Final Call"*, the plaintiff successfully sued a newspaper for libel over its unauthorized publication of a photograph of her which newspaper staff had randomly selected and doctored to make it appear that she was a criminal dressed in prison attire. 764 N.Y.S.2d 416, 417–18, 420 (N.Y. App. Div. 2003).

supporting role in disputes they did not cause.[40] In those cases, the alleged deepfake will not be what triggered the lawsuit or indictment. Rather, the deepfake will be just another piece of evidence in the course of litigation, where video evidence is already common.[41] In pre-trial and trial practice, deepfakes will touch every role in the courtroom: lawyers attempting to introduce or exclude videos as evidence; judges determining whether a video is admissible; expert and lay witnesses asked to testify about the video, and, finally; jurors weighing the evidence in order to reach a verdict.[42]

[40] For instance, in a child custody battle in the United Kingdom last year, the mother was revealed to have entered a doctored audio file into evidence. To support her contention that the father was too violent to be allowed access to their children, she had "used software and online tutorials to put together a plausible audio file" that sounded like a recording of him threatening her on a phone call. After the father's counsel obtained the original audio file, studied its metadata, and exposed the ruse, the court reportedly dismissed the fake evidence. Patrick Ryan, *'Deepfake' Audio Evidence Used in UK Court to Discredit Dubai Dad*, THE NATIONAL (Feb. 8, 2020), https://www.thenational.ae/uae/courts/deepfake-audio-evidence-used-in-uk-court-to-discredit-dubai-dad-1.975764. While the story refers to the file as a "deepfake," it's not clear whether the software the mother used was an AI-based synthetic audio app or something less sophisticated. *See* Wiggers, *supra* note 14.

[41] With respect to criminal cases, for example, "juries expect video to be presented to them in every case, whether it exists or not," which is understandable given that "some [commentators] estimate that video evidence is involved in about 80 percent of crimes." *See Video Evidence: A Primer for Prosecutors* 1, 2, U.S. DEP'T OF JUST., BUREAU OF JUST. ASSISTANCE (October 2016), https://it.ojp.gov/GIST/1194/File/FINAL-Video-Evidence-Primer-for-Prosecutors.pdf/ (citing Dale Garrison, *Advanced Video Forensics*, EVIDENCE TECH. MAG., July–August 2014 Issue, www.evidencemagazine.com/index.php?option=com_content&task=view&id=1688&Itemid=49).

[42] Unsurprisingly, this topic has drawn more attention from practitioners than from academics. I credit Chicago attorney Jonathan Mraunac as the first person to have published anything about the evidentiary issues that deepfakes raise for the courts. Jonathan Mraunac, *The Future of Authenticating Audio and Video Evidence*, LAW360 (July 26, 2018), https://www.law360.com/articles/1067033/the-future-of-authenticating-audio-and-video-evidence. Since then, other practitioners have also covered this topic. *E.g.*, Theodore F. Claypoole, *AI and Evidence: Let's Start to Worry*, THE NAT'L L. REV. (Nov. 14, 2019), https://www.natlawreview.com/article/ai-and-evidence-let-s-start-to-worry; Ashley Dean, *Deepfakes, Pose Detection, and the Death of "Seeing Is Believing"*, L. TECH. TODAY (July 10, 2019), https://www.lawtechnologytoday.org/2019/07/deepfakes-pose-detection-and-the-death-of-seeing-is-believing/; Kathryn Lehman, Scott Edson & Victoria Smith, "5 Ways to Confront Potential Deepfake Evidence in Court," LAW360 (July 26, 2019), https://www.law360.com/articles/1181306/5-ways-to-confront-potential-deepfake-evidence-in-court; Marie-Helen Maras & Alex Alexandrou, *Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and in the Wake of Deepfake Videos*, 23(3) INT'L J. EVID. & PROOF 255–62 (July 2019); Jason Tashea, *As Deepfakes Make It Harder to Discern Truth, Lawyers Can Be Gatekeepers*, ABA JOURNAL LAW SCRIBBLER (Feb. 26, 2019), http://www.abajournal.com/lawscribbler/article/as-deepfakes-make-it-harder-to-discern-truth-lawyers-can-be-gatekeepers.

Deepfakes may arise in the evidentiary context in a number of ways. A party might fabricate a video specifically for purposes of the litigation in order to try to prevail.[43] Or a litigant might encounter a deepfake video made by someone else and wish to introduce it into evidence, not realizing it is fake. Fake videos may end up (whether accidentally or maliciously) in archives that have historically been considered trustworthy, such as those of news outlets.[44] If their presence goes undetected by the custodian of those records, there is a risk that the custodian might unwittingly vouch for a deepfake when called upon to authenticate evidence in a court proceeding.

What is more, even in cases that do not involve fake videos, the very existence of deepfakes will complicate the task of authenticating *real* evidence. The opponent of an authentic video may allege that it is a deepfake in order to try to exclude it from evidence or at least sow doubt in the jury's minds.[45] Eventually, courts may see a "reverse *CSI* effect" among jurors.[46] In the age of deepfakes, jurors may start expecting the proponent of a video to use sophisticated technology to prove to their satisfaction that the video is *not* fake. More broadly, if juries—entrusted with the crucial role of finders of fact—start to doubt that it is possible to know what is real, their skepticism could undermine the justice system as a whole.

III. WHAT, ME WORRY?

As a novel technology that has obvious abusive applications while also being readily accessible to laypeople, deepfakes have already begun "striking fear" in observers that they "will spell the end of truth . . . as we know it" for our entire society.[47] For Jordan Peele, he of the fake Obama video, the deepfake medium *was* the message: a warning about the dystopian dangers that deepfake videos

[43] *E.g.*, Ryan, *supra* note 40 (describing doctored audio file used as evidence in child custody battle).

[44] Trust in the news media has been declining in America. Megan Brenan, *Americans' Trust in Mass Media Edges Down to 41%*, GALLUP (Sept. 26, 2019), https://news.gallup.com/poll/267047/americans-trust-mass-media-edges-down.aspx. Were news outlets to begin erroneously publishing videos that were later exposed as deepfakes, even more people would stop trusting the media, whether they believed the publication was accidental (*i.e.*, news outlets do not report the facts accurately) or intentional (*i.e.*, news outlets are trying to dupe the public).

[45] United States v. Tin Yat Chin, 371 F.3d 31, 38 (2d Cir. 2004) (once evidence is admitted, "'the evidence's persuasive force is left to the jury.' . . . [T]he other party then remains free to challenge the reliability of the evidence, minimize its importance, or to argue alternative interpretations of its meaning . . . . ") (quoting United States v. Dhinsa, 243 F.3d 635, 658 (2d Cir. 2001)).

[46] *See infra* Section V.C.

[47] Jeffrey Westling, *Deep Fakes: Let's Not Go Off the Deep End*, TECHDIRT (Jan. 30, 2019), https://www.techdirt.com/articles/20190128/13215341478/deep-fakes-lets-not-go-off-deep-end.shtml.

may pose if an overly credulous American public believes everything we see in the media or online.[48]

That fear is understandable. Seeing is believing. People tend to accept images "at face value."[49] Thanks to the probative value we attach to photos, "[a] photograph passes for incontrovertible proof that a given thing happened."[50] This assumption leaves people susceptible to being misled, because they will be convinced "irrespective of whether the videos and images have been fabricated."[51] If the software tools for creating deepfakes remain widely available and continue improving in quality of output, laypeople soon (if not already) will no longer be able to tell that a particular video is fake.

Nevertheless, there are historical reasons to doubt that deepfakes herald the death of truth. Deepfakes are just the latest chapter in the long history of fakery. Every kind of document, including imagery, is susceptible to manipulation. Society is aware of that possibility and has adjusted accordingly. From photography's earliest days, the public recognized that photographs could depict reality but were equally capable of illusion.[52] More recently, in the 1990s, pundits predicted that the introduction of Adobe Photoshop would precipitate a "crisis of truth"—but "society caught on and adapted to the technology."[53] Indeed, in 2004, a Photoshopped image of then-presidential candidate John Kerry went viral online; but as the internet corrected itself, search engine results for the doctored photo soon elevated hoax explanations above the photo itself.[54]

The courts, too, are no stranger to doctored photographs. "Modern jurors have been raised to believe that the camera does not lie, but they also have been exposed to the possibility that a camera can be made to lie."[55] Indeed, faked photos have been getting debunked in U.S. courts for 150 years: in an 1869 fraud

[48] David Mack, *This PSA About Fake News From Barack Obama Is Not What It Appears*, BUZZFEED NEWS (Apr. 17, 2018), https://www.buzzfeednews.com/article/davidmack/obama-fake-news-jordan-peele-psa-video-buzzfeed (describing Peele's urging in the video that viewers "'stay woke' by being vigilant to media sources" lest disinformation lead the country into dystopia).

[49] Maras & Alexandrou, *supra* note 42, at 257 (citations omitted).

[50] Sontag, *supra* note 1, at 3.

[51] Maras & Alexandrou, *supra* note 42, at 257 (citations omitted).

[52] Berger, *supra* note 32, at 183–84 ("The fact that photographs could 'lie' should not have been a surprise to most people familiar with photography in 1869 . . . . The public would have been well aware" of photography's capacity for sophisticated manipulation.) (citations omitted). Berger describes the mid-nineteenth century American public's attitude toward the nascent medium of photography as one of ambivalence: "it was simultaneously associated with science and magic, with commerce and art, with fraudulence and truthfulness . . . ." *Id.* at 182.

[53] Westling, *supra* note 47.

[54] Ken Light, *Fonda, Kerry and Photo Fakery*, WASH. POST (Feb. 28, 2004), https://www.washingtonpost.com/archive/opinions/2004/02/28/fonda-kerry-and-photo-fakery/15bdc6ed-c568-49fc-bddd-ac534c426865/.

[55] Lehman, Edson, & Smith, *supra* note 42.

case against a "spirit photographer" (who claimed he could "photograph the ghosts of the departed"), members of a professional association for photographers took the stand to attest to "the manipulability of photographs."[56] They held themselves out as experts deserving of credibility regarding the young medium: "as professional photographers *they* were the ones especially adept at the task of detection."[57] Drawing on that professed expertise, they pointed out the telltale signs of manipulation in the defendant's photographs admitted into evidence, such as how the shadow cast by the living person in the photo pointed in a different direction than that of the "ghost."[58]

As quaint as spirit photography might seem to us now, the case reminds us that the courts are accustomed to the possibility that a document is not what it is purported to be. The courtroom is a microcosm of society in general, but with more formal guardrails in place for validating the evidence that passes through it. To assure at least some baseline likelihood that an item of evidence "is what the proponent claims it is,"[59] courts have long imposed authentication requirements on those items, be they handwritten[60] or typed documents,[61] film[62]

[56] Berger, *supra* note 32, at 9–10, 185.

[57] *Id.* (citing Jennifer L. Mnookin, *The Image of Truth: Photographic Evidence and the Power of Analogy*, 10 YALE J. L. & HUMANITIES 1, 38 (1998)). As Berger describes it, Mnookin's article cast spirit photography "as integral to the judicial construction of photographs as evidence." *Id.* at 126.

[58] *Id.* at 185.

[59] FED. R. EVID. 901(a).

[60] *E.g.*, West v. State, 22 N.J.L. 212, 241–42 (1849) (in criminal forgery case, stating the common-law rule to be that authenticating a handwritten document required the testimony of "a witness having proper knowledge of the party's handwriting," or, failing that, expert testimony based on either prior "knowledge of the handwriting" or comparing the document in question to "other documents admitted to be genuine, or proved to have been treated and acted upon as such") (citations omitted).

[61] Prior to the Federal Rules of Evidence, "[t]he generally accepted rule" in the federal courts was "to the effect that the mere fact that a letter (other than a reply letter) purports to have been written and signed by the person in question is insufficient to establish its authenticity and genuineness," especially "where the letter is typewritten or printed . . . ." Nicola v. United States, 72 F.2d 780, 782 (3d Cir. 1934) (quoting 9 A.L.R. 987, 988 (1920)). "In order to make it evidence, it must be shown either to have been written by the person against whom it is produced, or by some one authorized to act in his behalf." *Id.* at 782–83.

[62] *E.g.*, Goldsboro v. Cent. R. Co., 37 A. 433, 434 (N.J. 1897) (photographs "are not admissible unless authenticated by other evidence that they are correct resemblances or truthful representations") (citations omitted); Miller v. Dumon, 64 P. 804, 805 (Wash. 1901) ("Photographs taken by the common processes are generally held admissible as evidence . . . when verified by proof that [the photograph] is a true representation of an object which is the subject of inquiry.") (citations omitted).

or digital photographs,[63] films, videotapes,[64] or digital videos.[65] If a document cannot be satisfactorily authenticated by its proponent, it will not be admitted into evidence.[66]

In short, generations of technologies with truth-subversive potential have become commonplace in society over the years. While the resulting fakes have inevitably gained traction at times in the public consciousness, the sky has not fallen. Past experience suggests that "the panic around" deepfakes will turn out to be a "false alarm," as one commentator recently concluded.[67] As far as the courts are concerned, I predict that this panic will prove largely unfounded. The nation's courts are robust institutions that have shown themselves capable of handling each new variant of the age-old problem of fakery. Courts' track record of resilience should assuage some of the hysteria that has crept into the discourse around deepfake technology.

[63] *E.g.*, Lorraine v. Markel Am. Ins. Co., 241 F.R.D. 534, 561 (D. Md. 2007) (noting that "[d]igital photographs present unique authentication problems because they are a form of electronically produced evidence that may be manipulated and altered," before going on to state that "[a]n original digital photograph may be authenticated the same way as a film photo, by a witness with personal knowledge of the scene depicted who can testify that the photo fairly and accurately depicts it.") (citing Edward J. Imwinkelried, *Can this Photo be Trusted?*, TRIAL, *49 (Oct. 2005), https://www.thefreelibrary.com/Can+this+photo+be+trusted%3F+Digital+photos+can+be+enhanced+to+help. . .-a0137876592.).

[64] *E.g.*, People v. Bowley, 382 P.2d 591, 594, 596 (Cal. 1963) (films can be authenticated by "the testimony of a person who was present at the time a film was made that it accurately depicts what it purports to show," or "by the aid of expert testimony . . . although there is no one qualified to authenticate it from personal observation"); State v. Newman, 484 P.2d 473, 477 (Wash. Ct. App. 1971) ("The requirements for the admission of video tapes should be similar to those for photographs," for which "it is only required that some witness, not necessarily the photographer, be able to give some indication as to when, where, and under what circumstances the photograph was taken, and that the photograph accurately portrays the subject illustrated") (citations omitted).

[65] *E.g.*, Commonwealth v. Rogers, 945 N.E.2d 295, 311 (Mass. 2011) (holding that a digital video recording was properly admitted where "the digital video recording was authenticated by the officer who created the compilation," who "described the process used to create it," and "[t]he judge took the precaution of conducting a voir dire before determining that the recording fairly and accurately presented what it purported to be").

[66] FED. R. EVID. 901(a).

[67] Russell Brandom, *Deepfake Propaganda is not a Real Problem*, THE VERGE (Mar. 15, 2019), https://www.theverge.com/2019/3/5/18251736/deepfake-propaganda-misinformation-troll-video-hoax ("We've had the tools to fabricate videos and photos for a long time. . . . AI tools can make that process easier and more accessible, but it's easy and accessible already. . . . [D]eepfakes are already in reach for anyone who wants to cause trouble on the internet. It's not that the tech isn't ready yet. It just isn't useful.") (citing Light, *supra* note 54).

IV. DEEPFAKES' POTENTIAL RAMIFICATIONS FOR TRIAL PRACTICE

This section reviews the basics of video evidence authentication, then looks at potential strategies for keeping deepfake videos out of evidence. I conclude that deepfake technology does not warrant changing the rules for authentication of video evidence to be more restrictive. The section then looks at deepfakes' potential impact on *real* evidence, ending with a cautionary note about access to justice.

A. Authentication Standards for Video Evidence

Authentication is fundamental to the admissibility of evidence.[68] "The bar for authentication of evidence is not particularly high."[69] Generally, the authentication requirement is satisfied by "evidence sufficient to support a finding that the item is what the proponent claims it is."[70] The proponent "need only make a prima facie showing of authenticity 'so that a reasonable juror could find in favor of authenticity or identification.'"[71] Once the threshold requirement of authentication is satisfied, "[t]he ultimate determination as to whether the evidence is, in fact, what its proponent claims is thereafter a matter for the jury."[72]

Courts treat video authentication as they do photographs. Both are "typically authenticated by showing [they are] a fair and accurate representation of the scene depicted." [73] The witness may—but need not—be the person who took the photo or video; the witness can also be someone who saw the event being recorded,[74] or who is otherwise "able to give some indication as to when, where, and under what circumstances the [video] was taken, and that the [video] accurately portrays the subject illustrated."[75]

Firsthand knowledge by the authenticating witness of the events depicted is preferable, but not required.[76] Despite not having been "present at the recording

[68] In addition to authentication, in order for a piece of evidence to be admissible, "[o]f course it must also be relevant to an issue at trial." United States v. Cejas, 761 F.3d 717, 723 (7th Cir. 2014) (citing FED. R. EVID. 401, 402).

[69] United States v. Gagliardi, 506 F.3d 140, 151 (2d Cir. 2007) (citing United States v. Dhinsa, 243 F.3d 635, 658 (2d Cir. 2001)).

[70] FED. R. EVID. 901(a).

[71] United States v. Workinger, 90 F.3d 1409, 1415 (9th Cir. 1996) (citation omitted).

[72] United States v. Vayner, 769 F.3d 125, 130 (2d Cir. 2014). "[T]he opposing party 'remains free to challenge the reliability of the evidence, minimize its importance, or to argue alternative interpretations of its meaning, but these and similar other challenges go to the *weight* of the evidence—not its *admissibility*.'" *Id.* at 131 (quoting United States v. Tin Yat Chin, 371 F.3d 31, 38 (2d Cir. 2004)).

[73] People v. Goldsmith, 326 P.3d 239, 245 (Cal. 2014) (citations omitted).

[74] *Id*.

[75] State v. Newman, 484 P.2d 473, 477 (Wash. Ct. App. 1971).

[76] State v. Sapp, 332 P.3d 1058, 1061 (Wash. Ct. App. 2014) (disapproving of the contrary holding in Saldivar v. Momah, 186 P.3d 1117, 1135 (Wash. Ct. App. 2008)). *See also*

of the exhibit . . . . A witness with prior knowledge of the people and places depicted in the exhibit could still establish when the exhibit was created based on the age of people in the exhibit or things depicted in the background."[77] Thus, a video can be authenticated by a witness testifying along the lines of, "I recognize the person speaking as the defendant, that's how he looked during the time period at issue, and that's his voice."

B. Authentication Challenges to Deepfakes

With a deepfake, of course, there can be no firsthand witness to the video's "recording." A liberal policy for authentication of photo and video evidence thus leaves room for a witness familiar with the person depicted to unwittingly vouch for a forgery by identifying the person's face and voice.

The risk could be even greater in courts that recognize the "silent witness" theory of video authentication, which "focuses on the automatic operation of the recording device and does not consider a witness's observations of the recorded events because the recording speaks for itself."[78] Another risk lies in materials held in third-party archives that historically have been considered trustworthy. Examples include a newsroom's archives (which are likely to hold both footage recorded by its own staff and cellphone videos contributed by eyewitnesses) or government databases (which nowadays include police department databases of officers' body camera footage). Records held in such archives are often considered self-authenticating,[79] or require only that a custodian of records certify copies of them.[80]

In this age of hacking and data breaches, records archives need to worry about their cybersecurity, not just physical chain of custody.[81] In 2018, a researcher demonstrated that for many police body cameras then on the market, their videos

Goldsmith, 326 P.3d at 245 (setting forth the multiple permissible means of authentication besides eyewitness testimony, including "other witness testimony, circumstantial evidence, content and location," or, pursuant to state statute, "'any other means provided by law,' including a statutory presumption") (citations omitted).

[77] *Sapp*, 332 P.3d at 1061.

[78] Mraunac, *supra* note 42.

[79] *See, e.g.*, United States v. Loera, No. 09-cr-0466, 2018 U.S. Dist. LEXIS 96132, at *11 (E.D.N.Y. June 7, 2018) (holding, in a case against notorious drug cartel kingpin, "El Chapo," that video evidence consisting of news footage was self-authenticating under Rule 902(6) of the Federal Rules of Evidence and commenting that "[i]t would be extremely difficult to forge news videos") (citations omitted).

[80] *E.g.*, FED. R. EVID. 902(4)(A) ("A copy of an official record" is self-authenticating "if the copy is certified as correct by . . . the custodian or another person authorized to make the certification").

[81] Data breaches increased by 17 percent from 2018 to 2019, with hacking being "the most common form of data breaches." Chris Morris, *Hackers Had a Banner Year in 2019*, FORTUNE (Jan. 28, 2020), https://fortune.com/2020/01/28/2019-data-breach-increases-hackers/.

could be remotely downloaded, digitally manipulated, and re-uploaded.[82] Further, "[t]he bodycams don't have a cryptographic mechanism to confirm the validity of the video files they record either," meaning that "when the devices sync with a cloud server or station PC, there's no way to guarantee that the footage coming off the camera is intact."[83] That leaves room for manipulated footage to be uploaded to a body camera, and then, when the camera is synced to the cloud, to infiltrate the police database. If the integrity of a video in a database could be compromised without the knowledge of the witness who is called upon to authenticate it, then the witness (for example an archive's custodian) could unwittingly offer inaccurate testimony about an altered or ersatz video.

The possibility of remote tampering may undermine the reliability of video footage in third-party databases.[84] As such, the proponent of such evidence may be required to do more to establish the video's authenticity. Even in silent-witness jurisdictions, "many courts still require some additional evidence or testimony providing a 'strong showing of authenticity and competency, including proof that the evidence was not altered.'"[85] Thus, "[a] trial judge should consider, among other things, 'any evidence of editing or tampering' before admitting a recording under the silent witness theory."[86]

To challenge a suspected deepfake video's authenticity, the opposing party could move to strike (in a civil case)[87] or exclude (in a criminal case)[88] the video, and produce some evidence in support of the contention that it is doctored or fake.[89] For example, the officer who wore a body camera might testify that the

[82] Lily Hay Newman, *Police Bodycams Can Be Hacked to Doctor Footage*, WIRED (Aug. 11, 2018), https://www.wired.com/story/police-body-camera-vulnerabilities/. The researcher found that in four out of five body camera models tested, "vulnerabilities would allow an attacker to download footage off a camera, edit things out or potentially make more intricate modifications, and then re-upload it, leaving no indication of the change." *Id.*

[83] *Id.* In response to the possibility of remote tampering, Amber, a company that makes authentication technology for police body-worn cameras, brought that researcher on as "an adviser on cybersecurity threats facing police body cams." *Ticks*, *supra* note 25, at 43.

[84] *Ticks*, *supra* note 25, at 43 (If archives "are, or become, vulnerable, then the verification of media they provide will essentially become meaningless.").

[85] Lehman, Edson, & Smith, *supra* note 42 (footnote omitted).

[86] *Id.* (footnote omitted).

[87] At the summary judgment stage, for example, "[a] motion to strike is the proper vehicle for challenging the admissibility of materials submitted in support" of the summary judgment motion. Goguen v. Textron, Inc., 234 F.R.D. 13, 16 (D. Mass. 2006) (citing 11 JAMES W. MOORE ET AL., MOORE'S FEDERAL PRACTICE § 56.14[4][a] (3d ed. 1997)). *See also* Lehman, Edson, & Smith, *supra* note 42 (providing detailed suggestions for deploying motion practice in order to confront potential deepfake evidence).

[88] *E.g.*, State v. Smith, 192 So.3d 836, 837 (La. Ct. App. 2016) (reviewing the denial of a motion to exclude digital evidence for lack of authentication by the state).

[89] *See* Webb v. Scott, No. 11-cv-128, 2014 U.S. Dist. LEXIS 63243, at *14 (D. Utah Mar. 14, 2014) (rejecting plaintiff's motion to have police dash camera video authenticated by an

footage did not match the officer's memory and identify the discrepancies. The burden would then shift to the party moving for the video's admission to present evidence sufficient for a reasonable juror to conclude that the evidence is what it purports to be.[90]

Likely the simplest way to challenge a deepfake video is for the person depicted in it (if available) to testify under oath that the video is bogus. This strategy, however, may be risky for individuals who are likely to have credibility problems even if their testimony is truthful, or (in criminal cases) for criminal defendants, who have the right not to testify.[91] If the witness does testify, opposing counsel might impeach the witness's credibility,[92] or jurors might make a negative assessment based on their own observations.[93] It will be up to the attorney challenging the video's authenticity to make the tactical decision whether calling that witness might do more harm than good.[94]

In addition to lay-witness testimony, litigators should be able to exclude deepfakes from evidence through existing strategies for challenging a video's provenance and chain of custody, including: targeted discovery,[95] using cross-examination to grill a lay witness called to vouch for the video's authenticity,[96] and calling experts trained in digital video forensics to testify.[97] With the development of AI-powered deepfake-detection systems,[98] digital forensics

expert, because "as the proponent of the claim about evidence tampering, Plaintiff bears the burden to prove the claim in these civil proceedings."). The proponent of a third-party document cannot shift the burden of authenticating it to the opposing party. Adobe Sys. v. Christenson, No. 10-cv-422, 2011 U.S. Dist. LEXIS 16977, at *28 (D. Nev. Feb. 7, 2011) (citing FDIC v. Halpern, 271 F.R.D. 191 (D. Nev. 2010)).

[90] *E.g.*, *Smith*, 192 So.3d at 842-43.

[91] *See* U.S. CONST. amend. V.

[92] FED. R. EVID. 608(a) ("A witness's credibility may be attacked . . . by testimony about the witness's reputation for having a character for truthfulness or untruthfulness, or by testimony in the form of an opinion about that character."). There is added risk if the witness has a criminal record: in certain circumstances, witnesses can be impeached by evidence of past criminal convictions. FED. R. EVID. 609.

[93] Jurors assess a witness's credibility based on their "evaluations of a witness' demeanor, perception, memory, narration and sincerity." Steven I. Friedland, *On Common Sense and the Evaluation of Witness Credibility*, 40 CASE W. RES. L. REV. 165, 174 (1989). These assessments are likely to be negative "if a [witness] has memory problems, cognitive or mental-health issues, or just lacks experience speaking to a group, [because] that person will have a tough time explaining himself and dealing with cross-examination." Toni Messina, *Why Defendants Rarely Testify*, ABOVE THE LAW (Sept. 30, 2019), https://abovethelaw.com/2019/09/why-defendants-rarely-testify/.

[94] *See Gonzalez v. United States*, 553 U.S. 242, 249 (2008) (tactical decisions for counsel to make in criminal trial management include "the witnesses to call").

[95] Lehman, Edson, & Smith, *supra* note 42.

[96] *Id.*

[97] *Ticks*, *supra* note 25, at 29.

[98] *See supra* note 22 and accompanying discussion.

experts are poised to play a significant role in court authentication battles over suspected deepfakes.[99] The modern-day colleagues of the professional photographers who testified in that 1869 spirit photographer case[100] are “forensic video analysts, who are in the business of detecting fake videos and images for criminal and civil courts.”[101] Grant Fredericks, the president of Forensic Video Solutions and a “pioneer in the field,” is confident that fake videos will be kept out of evidence, both because they can be readily discovered using the advanced tools of his trade and because the video’s proponent would be unable to answer basic questions to authenticate it (who created the video, when, and with what technology).[102]

Sophisticated forensics tools won’t be necessary in every single case, at least not yet.[103] That’s because at present, “‘[t]he quality of many deepfake generated videos makes it relatively easy to detect a manipulation without requiring an extensive forensic investigation.’”[104] For example, in a recent U.K. court case involving a doctored audio file made by one of the parties, analysis of the file’s metadata enabled opposing counsel to uncover the deception.[105] And in May 2019, a video circulating on social media of House Speaker Nancy Pelosi, which seemed to show her drunkenly slurring her words, was easily exposed as having been simply altered to slow down the audio—hardly a high-tech “deepfake.”[106] If many manipulated videos are poor-quality and easy to spot, then it stands to reason that in many court cases, excluding them will not be a heavy lift.

[99] Mraunac, *supra* note 42 (“Naturally, when a party questions the authenticity of video evidence, expert testimony becomes relevant . . . . If additional evidence-authentication requirements develop, [expert witnesses] should expect their law-firm and corporate clients to rely even more heavily on their technological expertise than they do currently.”).

[100] *See supra* notes 56–58 and accompanying text.

[101] Mark J. Pescatore, *Forensic Video Experts: Fake Videos Not Threat to Courtroom Evidence*, PIPELINE COMM. (June 24, 2019), https://www.pipecomm.com/2019/06/24/forensic-video-experts-fake-videos-not-threat-to-courtroom-evidence/.

[102] *Id.* It bears noting—given that this whole Article is basically about media literacy—that this story was written by a public relations firm, not a journalistic outlet.

[103] Tashea, *supra* note 42.

[104] *Id.* (quoting Matt Turek, program manager at DARPA).

[105] Ryan, *supra* note 39. The article is unclear whether the opposing counsel enlisted an expert to assist in the metadata analysis.

[106] Laura Hazard Owen, *What Do We Do About the “Shallowfake” Nancy Pelosi Video and Others Like It?*, NIEMAN LAB (May 31, 2019), https://www.niemanlab.org/2019/05/what-do-we-do-about-the-shallowfake-nancy-pelosi-video-and-others-like-it/. Such low-tech doctored videos are typically called “shallowfakes” or “cheapfakes” to contrast them against sophisticated AI-driven deepfakes. Tony Romm, Drew Harwell, and Isaac Stanley-Becker, *Facebook Bans Deepfakes, but New Policy May Not Cover Controversial Pelosi Video*, WASH. POST (Jan. 7, 2020), https://www.washingtonpost.com/technology/2020/01/06/facebook-ban-deepfakes-sources-say-new-policy-may-not-cover-controversial-pelosi-video/ (showing original and altered videos side-by-side).

With ongoing advances in deepfake technology, however, that will not be the case for long.[107] Soon, "'detecting the manipulations will require more sophisticated technologies and forensic techniques.'"[108] Even now, a bad actor could be willing to invest the time, money, and computing power necessary to make a high-quality deepfake, if the potential payoff is high enough.[109] Once such situations arise in court, the ability of an expert witness to apply deepfake-detection and video-authentication tools to the video in question will be necessary to bar fake evidence from infecting the proceedings.[110] That will require both the expert and the tools the expert uses to survive challenges by opposing counsel as to the expert's qualifications and to the validity of the tools on which the expert's testimony is based.[111]

It may take time for video-authentication tools now on the market, or deepfake-detection techniques being developed in academic and corporate

[107] William A. Galston, *Is Seeing Still Believing? The Deepfake Challenge to Truth in Politics*, BROOKINGS INST. (Jan. 8, 2020), https://www.brookings.edu/research/is-seeing-still-believing-the-deepfake-challenge-to-truth-in-politics/ ("The capacity to generate deepfakes is proceeding much faster than the ability to detect them.").

[108] Tashea, *supra* note 42 (quoting Matt Turek, Program Manager at DARPA).

[109] The current method of making "[m]ost of the deepfakes that are shared online . . . can take hours or days even with access to expensive hardware, and even longer with consumer-grade PC components." Samantha Cole, *This Program Makes It Even Easier to Make Deepfakes*, VICE MOTHERBOARD (Aug. 19, 2019), https://www.vice.com/en_us/article/kz4amx/fsgan-program-makes-it-even-easier-to-make-deepfakes. Cutting-edge deepfake programs also aren't "cheap or easy to make": the FSGAN "face-swapping" program made by Israeli university researchers "required eight Nvidia Tesla v100 GPU processors—which can cost around $10,000 each for consumers—to train the generative adversarial network that the program then uses to create deepfakes in real-time." *Id*.

[110] *See* Pescatore, *supra* note101, and text accompanying note 102. There is some question whether those expert witnesses will be able to succeed: "as machine learning and AI technology advances, the testimony of digital media forensic experts may not be enough to authenticate evidence, because even expert witnesses may not be able to discern the modifications made to digital videos." Maras & Alexandrou, *supra* note 42, at 259.

[111] The tests for the admissibility of expert testimony and for experts' new scientific tools, techniques, and methodologies vary by jurisdiction, and are in flux in some states. The federal courts and a majority of state courts follow the standard initially articulated in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 597 (1993), and while a minority of states still employ the older test set forth in *Frye v. United States*, 293 F.1013, 1014 (D.C. Cir. 1923), some are warming up to *Daubert*. *See generally* Sean M. McDonough, *The* Daubert *Standard Once Again Controls in Florida State Court*, NAT'L L. REV. (June 5, 2019), https://www.natlawreview.com/article/daubert-standard-once-again-controls-florida-state-court (discussing Florida Supreme Court's 2019 switch from *Frye* to *Daubert*). *See also* David L. Faigman & Edward J. Imwinkelreid, *Wading into the* Daubert *Tide:* Sargon Enterprises, Inc. v. University of Southern California, 64 HASTINGS L.J. 1665, 1665 (2013) (discussing California Supreme Court's 2012 decision in *Sargon* to take steps toward, but not fully adopt, the *Daubert* test).

research labs, to be accepted by the courts as a basis for expert testimony.[112] Nonetheless, the deepfakes arms race is sure to spawn a cottage industry, albeit a modestly-sized one, of expert witnesses who can assess disputed videos. At present, there are likely only a handful of people who are qualified to do so. The necessary forensic expertise to detect deepfakes requires a specialized background in a field with complex barriers to entry, meaning the few who have it can command top dollar for their services.[113] What is more, someone who is qualified to testify as an expert in some other area of media forensics will not necessarily pass muster as an expert in the domain of deepfake detection.[114] "In this context, the expert witness for audio and video authentication would no longer be an acoustical engineer or visual image expert but a software engineer, cryptographer and/or a representative from the hardware manufacturer."[115] Therefore, the number of experts qualified to evaluate whether or not a video is a deepfake will not grow quickly. Qualified experts will thus be able to charge a premium for their service as expert witnesses in court.

Does the potential for faking digital video warrant raising the current bar for authentication of digital video evidence? Courts have disagreed. Ten years ago, in *People v. Beckley*, a California appellate court ruled that a photograph downloaded from social media website MySpace had not been properly authenticated because there was no testimony either from any witness who was present when the photo was taken or from any expert who could say the photo was *not* faked.[116] "Such expert testimony," the court wrote, is critical "to prevent the admission of manipulated images," given that "digital photographs can be changed to produce false images" without the need for much skill, thanks to "the advent of computer software programs such as Adobe Photoshop."[117]

[112] As Chief Justice John Roberts has acknowledged, the courts are slow to embrace new technologies. Stephanie Condon, *John Roberts: Courts Will Always Be Slow to Embrace "The Next Big Thing"*, CBS NEWS (Jan. 5, 2015), https://www.cbsnews.com/news/john-roberts-courts-will-always-be-slow-to-embrace-the-next-big-thing/.

[113] *Ticks*, *supra* note 25, at 29.

[114] Maras & Alexandrou, *supra* note 42, at 259 ("experts in the field of image and video forensics analysis must also, in the near future, be well-versed in machine learning and artificial intelligence as applied to the field of media forensics to explain why the results obtained using them are valid and reliable."). *See* Jones v. Union Pac. R. R. Co., No. 12-cv-771, 2015 U.S. Dist. LEXIS 118928, at *35–39 (N.D. Ill. Sept. 8, 2015) (finding that a witness was unqualified to offer expert testimony about video recordings' authenticity, where he was "not a licensed forensic analyst" and lacked the requisite "training, experience, or expertise that would enable him to perform a forensic analysis of the videos" at issue, even though he "may have experience in voice identification, audio and video editing and production, and perhaps even forensic *audio* analysis"). *See also Ticks*, *supra* note 25, at 8 ("The field of media forensics has only developed over the last two decades, and until recently, was still considered to be a niche field. Media forensics is not only a new field, but a disputed one.").

[115] Mraunac, *supra* note 42.

[116] 110 Cal. Rptr. 3d. 362, 366 (Cal. Ct. App. 2010).

[117] *Id.*

Another California appeals court case, *In re K.B.*, subsequently rejected *Beckley*'s approach as inconsistent with the California Supreme Court's restatement of the existing rule that eyewitness or photographer testimony "may, but need not," be what authenticates a photograph.[118] In short, the potential for manipulation does not justify narrowing the bases for authenticating a digital photograph. That view was recently echoed by a Colorado state appeals court in *People v. Gonzales*, which opined that while software has made it easy for laypeople to manipulate recordings, "the fact that the falsification of electronic recordings is always possible does not, in our view, justify restrictive rules of authentication that must be applied in every case when there is no colorable claim of alteration."[119]

These court cases, which span the last decade, place concerns over deepfake technology into context. While audio or video manipulation is a greater risk now than it used to be back before every Joe Average Computer Owner had the tools to do it,[120] that does not warrant reversing the current liberal authentication rules for video evidence and treating every digital video as suspect by default without some concrete reason. *Goldsmith*, *In re K.B.*, and *Gonzales*, decided between 2014 and 2019, show a continued commitment to that liberal policy, even as the technology for manipulating video has been growing ever *more* sophisticated than it was ten years ago when *Beckley* was decided.[121]

That indicates that courts are confident in the processes they already have in place for excluding manipulated evidence. I share that confidence. The protective processes that courts have developed over the years will, I predict, prove robust against deepfakes, as they have for previous generations of technology.[122] The existence of the mere possibility of manipulation, without

[118] In re K.B., 190 Cal. Rptr. 3d. 287, 293 (Cal. Ct. App. 2015) (quoting People v. Goldsmith, 326 P.3d 239, 245 (Cal. 2014)).

[119] People v. Gonzales, 2019 COA 30, ¶ 29 (citations omitted) (explaining "[w]hen a plausible claim of falsification is made by a party opposing the introduction of a recording, the court may and usually should apply additional scrutiny" to determine whether a reasonable jury could conclude that the item is what it purports to be).

[120] *Id.* at ¶ 28–29 ("There is no question that the alteration of electronic recordings, whether audio or video, is more of a risk today…") (citing Bruce E. Koenig & Douglas S. Lacey, *Forensic Authentication of Digital Audio and Video Files, in* HANDBOOK OF DIGITAL FORENSICS OF MULTIMEDIA DATA AND DEVICES 133 (Anthony T. S. Ho & Shujun Li eds., 2015)).

[121] *See generally* Brooke Borel, *Clicks, Lies, and Videotape*, SCI. AM. (Oct. 1, 2018), https://www.scientificamerican.com/article/clicks-lies-and-videotape/ (tracing evolution of AI-generated video from "advances in a type of AI called deep learning" in 2012, to "teach[ing] the AI to train itself" in 2014, to researchers' 2018 advancements of "'deep video,' which uses a type of GAN," and "figur[ing] out a way to get GANs to make incredibly high-resolution faces").

[122] Remember, image manipulation has been a known risk ever since the early years of photography when it involved "the manual manipulation of a negative during or after exposure." Berger, *supra* note 32, at 144. *See also* Joshua Rothman, *In the Age of A.I., Is*

more, does not call for a high bar for authentication today any more than it did 150 years ago.[123]

The current rules for authentication are adequate as-is. With that said, those rules do allow for disputes where there is some question of authenticity,[124] and those disputes might start to arise more often if deepfakes become more prevalent. We can foresee that evidentiary challenges to suspected deepfakes will add significantly to case timelines, and also "will likely increase the cost of litigation because new forensic techniques and expert witnesses aren't cheap."[125] Litigators will have to manage their clients' expectations accordingly.

C. Deepfakes' Ramifications for Genuine Evidence

Deepfakes' authentication difficulties are twofold. One problem is how to show a video is fake. The other is how to show it isn't. As deepfakes become increasingly common and realistic, their very existence will undermine the reliability of genuine evidence, creating headaches for the proponents of authentic videos.

In the era of digital evidence, practitioners are already accustomed to bringing and responding to courtroom challenges of digital photo, video, and sound recordings, despite Rule 901's relatively low standard for authentication.[126] But as real and fake become harder to distinguish, such challenges may be harder for

Seeing Still Believing?, THE NEW YORKER (Nov. 5, 2018), https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing ("the transposition, in a famous photograph from the eighteen-sixties, of Abraham Lincoln's head onto the body of the slavery advocate John C. Calhoun" was a "milestone[] in the history of image manipulation").

[123] The spirit photographer prosecuted for fraud in 1869 actually *won* his case, chastening the members of the professional association who had testified against him to the degree that when a different spirit photographer showed his pictures at an 1875 meeting of the association, one member cautiously withheld judgment, saying, "it is very easy to say a thing is a fraud, but it is quite another matter to prove it to be so." Berger, *supra* note 32, at 192–93. The spirit photographer agreed, stating that "it was hardly fair for gentlemen to call these pictures a fraud and a deception without they had better proofs than mere assertion." *Id.*

[124] *See* FED. R. EVID. 901.

[125] Tashea, *supra* note 42 (quoting TruePic legal officer Tara Vassefi); *see also Ticks*, *supra* note 25, at 29 ("Requiring additional verification for multimedia could lead to a more protracted and expensive legal process.").

[126] *See, e.g.*, Maya Leszczynski *et al.*, *Evidentiary Objections to Challenge Commonly Introduced Evidence Used in Support of Gang Allegations* 66–68, 71, Immigrant & Non-Citizen Rights Clinic, CUNY School of Law (July 2019), https://www.law.cuny.edu/wp-content/uploads/media-assets/Evidentiary-Objections_2019.pdf; Shawn McCann & Lauren Horwitz, *Laying the Foundation for Electronic and Documentary Evidence at Trial*, ADVOCATE (October 2015), https://www.advocatemagazine.com/article/2015-october/laying-the-foundation-for-electronic-and-documentary-evidence-at-trial-2; Amanda Hale, *Objecting to Video and Audio Evidence Without Hesitation*, JAMES PUBL'G (Oct. 7, 2013), https://jamespublishing.com/2013/objecting-video-audio-evidence-without-hesitation/; *supra* Section V.A and ensuing discussion.

the proponent to overcome. As described above, successfully getting a video admitted into evidence may require additional motion practice, witness testimony, and forensic tools.[127] These hurdles threaten to exclude *genuine* evidence from being admitted.

That risk is one of the motivators behind the video-authentication tools described above. As said, if deepfake technology evolves to the point where it becomes too difficult to prove the negative—that a video is *not* a fake—an alternative strategy is to prove affirmatively that a video is real, by attaching additional metadata to it at the moment the video is taken.[128] So-called "verified media capture technology" can help "to ensure that the evidence [users] are recording . . . is trusted and admissible to courts of law."[129] For example, an app called eyeWitness to Atrocities, "allows photos and videos to be captured with information that can firstly verify when and where the footage was taken, and secondly can confirm that the footage was not altered," all while the company's "transmission protocols and secure server system . . . create[] a chain of custody that allows this information to be presented in court."[130] That information, paired with the app-maker's willingness to provide a certification to the court or send a witness to testify if needed, could satisfy a court that the video is admissible, even if the videographer is unavailable.[131]

In federal court, recent changes in the evidence rules might simplify the process of getting videos admitted that were recorded using verified-capture tools. A pair of 2017 amendments to Federal Rule of Evidence 902 "were designed to simplify the legal process and reduce the costs associated with using electronically-stored information as evidence."[132] Upon proper certification, Rule 902(13) provides for self-authentication of "[a] record generated by an electronic process or system that produces an accurate result"; Rule 902(14), of "[d]ata copied from an electronic device, storage medium, or file, if authenticated by a process of digital identification."[133]

Tara Vassefi, then Washington Director of Strategic Initiatives at the photo and video verification platform Truepic, wrote in 2018 that these new rules should allow lawyers to establish quickly that video evidence is self-authenticating, without any need to call a witness (such as a custodian of records) to testify in person in court.[134] According to Vassefi, Truepic was designed to

[127] *See supra* Section V.B and ensuing discussion.

[128] *See supra* Section II.

[129] *Ticks*, *supra* note 25, at 22.

[130] *Id.* at 27 (quoting Wendy Betts, Project Director at eyeWitness to Atrocities).

[131] *Id.* at 29.

[132] *Id.* at 30.

[133] FED. R. EVID. 902(13), 902(14) (2017).

[134] Tara Vassefi, "A Law You've Never Heard of Could Help Protect Us From Deceptive Photos and Videos," UC Berkeley School of Law Human Rights Center (Nov. 30, 2018), https://medium.com/humanrightscenter/a-law-youve-never-heard-of-could-help-protect-us-from-fake-photos-and-videos-df07119aaeec.

meet these new evidentiary standards "by streamlining authentication for those with limited legal resources."[135] However, the version of the Truepic app on which Vassefi based her analysis was later "retired" in 2019.[136] At present, therefore, the question of video-authentication tools' ability to satisfy Rule 902's new self-authentication provisions remains unsettled. If that question is eventually answered in the affirmative, it may become simpler for genuine videos to be admitted into evidence cheaply and expeditiously, without prolonged authentication fights.

The battle over a video does not end after authentication succeeds and the video is admitted: the opposing party remains free to attack it. Admissibility and weight afforded by the fact-finder are two separate evidentiary issues.[137] After losing an authentication challenge, the opposing party may still attempt to persuade the jury to accord little weight to the video by questioning its reliability or downplaying its importance.[138] This is especially likely to happen where the video is highly damaging to the opposing party's case.

In order to minimize a video's impact on the jury, that party could attempt to capitalize on the jury's awareness that deepfakes exist.[139] If the media environment around us is rife with fakes, why should seeing equal believing? Why should the jury believe that this particular video is *not* a fake, just because it cleared the low bar for authentication? Awareness that deepfakes exist could put an end to the credibility that people have traditionally accorded to photo and video evidence.[140] As it becomes harder and harder to tell real and fake apart, "AI-manipulated digital videos may eventually have little (if any) probative value in courts."[141] That is positive, insofar as it helps keep factfinders from falling prey to deepfakes, but a reduction in video's probative value is undesirable when it comes to *real* evidence.

Professors Chesney and Citron call this phenomenon "the liar's dividend": "as the public becomes more educated about the threats posed by deep fakes," it will be more and more feasible for bad actors to "try to escape accountability for

[135] *Ticks*, *supra* note 25, at 30.

[136] Vassefi, *supra* note 134; Mounir Ibrahim, *Better Images. Better Decisions: Truepic's Quest to Positively Impact Society*, TRUEPIC (Nov. 2, 2019), https://medium.com/truepicinc/better-images-better-decisions-truepics-quest-to-positively-impact-society-900a82a2baa5 (Truepic "recently made the decision to retire the free mobile application of Truepic's technology, to be replaced at a later date.").

[137] United States v. Vayner, 769 F.3d 125, 131 (2d Cir. 2014).

[138] *Id.*

[139] *See* Galston, *supra* note 107 ("a year ago[,] . . . few would have understood what [a November 2019 *New York Times* headline containing the word 'deepfake'] meant. Today, most do.").

[140] *See* Maras & Alexandrou, *supra* note 42, at 257 ("People tend to believe what they see. For this reason, images and other forms of digital media are often accepted at face value. Digital images and videos are a powerful form of persuasion on a fact of a matter being asserted.") (citations omitted).

[141] *Id.* at 257–58 (citations omitted).

their actions by denouncing authentic video and audio as deep fakes."[142] Put simply: a skeptical public will be primed to doubt the authenticity of real audio and video evidence."[143]

Juries are the public in microcosm, so that skepticism will follow them into the jury box, where it will be ripe for exploitation by attorneys seeking to downplay damaging evidence.[144] Recall that even if the opponent of a video loses the battle over admissibility, that party may seek to minimize an authentic video's weight to the jury.[145] This approach may be especially tempting for criminal defense attorneys, given the "beyond a reasonable doubt" standard.[146] If the tactic is used successfully in enough cases, then as said, video evidence may lose some of the persuasive power it presently holds for people.

Indeed, there is a chance that litigators will start seeing a sort of "reverse *CSI* effect." The "*CSI* effect" refers to the phenomenon of jurors demanding high-tech evidence even in run-of-the-mill cases, thanks to the popular TV police procedural.[147] Similarly, the availability of software for authenticating real video, and of sophisticated AI tools for detecting deepfakes,[148] may have an unintended consequence. If jurors know such tools exist, they may accord little weight to a video unless the proponent either proves the positive—by showing

[142] Chesney & Citron*, supra* note 6 at 1785.

[143] *Id*. "The liar's dividend" has already had an effect in the real world, where video evidence at the center of recent controversies in Brazil, Gabon, and China was accused of being faked. "You can already see a material effect that deepfakes have had," said an engineer overseeing Google's deepfake research. "They have allowed people to claim that video evidence that would otherwise be very convincing is a fake." Metz, *supra* note 22.

[144] Chesney & Citron*, supra* note 6 at 1785.

[145] United States v. Vayner, 769 F.3d 125, 131 (2d Cir. 2014).

[146] The Fourteenth Amendment's "Due Process Clause protects the accused against conviction except upon proof beyond a reasonable doubt of every fact necessary to constitute the crime with which he is charged." *In re Winship*, 397 U.S. 358, 364 (1970); *see also Jackson v. Virginia*, 443 U.S. 307, 315–16 (1979) ("no person shall be made to suffer the onus of a criminal conviction except upon sufficient proof—defined as evidence necessary to convince a trier of fact beyond a reasonable doubt of the existence of every element of the offense").

[147] *See, e.g.*, Katie L. Dysart, *Managing the CSI Effect in Jurors*, AM. BAR ASS'N, SECTION OF LITIGATION, TRIAL EVID. CMTE. (May 28, 2012), https://www.americanbar.org/content/dam/aba/administrative/litigation/materials/2016_sac/written_materials/13_managing_the_csi_effect_in_jurors_tri.authcheckdam.pdf; Orin S. Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311, 349 (2012) (defining the "CSI effect" as the phenomenon "by which jurors in routine criminal cases expect prosecutors to introduce evidence collected using high-tech investigatory tools like those features on popular television dramas such as Law & Order and CSI"); Tom R. Tyler, *Viewing* CSI *and the Threshold of Guilt: Managing Truth and Justice in Reality and Fiction*, 115 YALE L.J. 1050, 1052 (2006) ("CSI effect" occurs when "people who watch the series develop unrealistic expectations about the type of evidence typically available during trials, which, in turn, increases the likelihood that they will have a 'reasonable doubt' about a defendant's guilt").

[148] *See supra* Section II.

the video was captured via a video-authentication tool and thus should be considered authentic[149]—or proves the negative, by using the latest detection technology (possibly at great expense) to satisfy the jury that the video is *not* a deepfake. Traditional means of persuading juries, such as the introduction of witness testimony to vouch for a video, may no longer work as well as they do now.

Taken far enough, jury skepticism about video evidence could eat away at public trust in the very institution of the courts. "As deep fakes become widespread," Chesney and Citron warn, "the public may have difficulty believing what their eyes or ears are telling them—even when the information is real."[150] If so, "the spread of deep fakes threatens to erode the trust necessary for democracy to function effectively."[151] The judicial branch is, of course, an indispensable part of a democracy.[152] If juries cease believing it is possible to discern what is true—or if they believe that court proceedings are riddled with fake evidence—then the courts will lose the public trust that they depend upon for their legitimacy.[153]

This gradual decay need not be the result of a malicious scheme by adversaries of democracy to undermine the courts. Rather, if it happens, it will likely be the culmination of a series of individual acts by trial lawyers who are motivated to win.[154] Challenges to what is in fact authentic evidence may be sincere, or they may be specious.[155] As discussed above, an authentication challenge can impose

[149] *See Ticks*, *supra* note 25, at 11–12 (describing "controlled-capture" and "verified-at-capture" tools for authenticating video, image, and audio recordings).

[150] Chesney & Citron, supra note 6, at 1786.

[151] *Id.*

[152] *E.g.*, Hon. Tassaduq Hussain Jillani, *Judicial Review and Democracy*, ABA LITIGATION JOURNAL (Jan. 1, 2018), https://www.americanbar.org/groups/litigation/publications/litigation_journal/2017-18/winter/judicial-review-and-democracy/ ("Democracy, as understood generally, is a political system . . . where the rule of law is enforced through an independent judiciary.").

[153] "Because the judicial branch relies heavily on public support to perform its role in our system of government, public trust and confidence is a precious commodity for the courts." *Public Trust and Confidence Resource Guide*, NAT'L CTR. FOR STATE COURTS, https://www.ncsc.org/Topics/Court-Community/Public-Trust-and-Confidence/Resource-Guide.aspx. *Cf.* United States v. Nixon, 418 U.S. 683, 709 (1974) ("The very integrity of the judicial system and public confidence in the system depend on full disclosure of all the facts, within the framework of the rules of evidence.").

[154] Attacking the validity and reliability of the opposing side's evidence is a standard, expected part of an attorney's role in the adversarial process. *Cf.* Strickland v. Washington, 466 U.S. 668, 685 (1984) (a fair trial involves "evidence subject to adversarial testing"). Attorneys have an "'overarching duty to advocate the defendant's cause'" via "legitimate, lawful conduct compatible with the very nature of a trial as a search for truth." Nix v. Whiteside, 475 U.S. 157, 166 (1986) (quoting *Strickland*, 466 U.S. at 688).

[155] Ethically, attorneys are barred only from making a "frivolous" argument, MODEL RULES OF PROF'L CONDUCT r. 3.1 (AM. BAR ASS'N 2019), meaning it is "both baseless and

significant additional costs on the proponent of a video.[156] Foreseeably, it will be a matter of strategic pre-trial practice, in some cases, for parties to claim a video (or image, or audio) is fake, particularly if it is really damaging evidence against that party.[157] If anticipating an authentication challenge, the proponent of a video may need to make a strategic decision about whether the video's probative value to the case outweighs the cost of getting it admitted. The expense and added delay might not be feasible or worth it to the client. This has access-to-justice ramifications. In slowing down the resolution of cases and increasing costs along the way, the additional burden of authenticating video threatens to "exclude[] those without access to experts and resources" from getting the justice they deserve.[158]

Consider a police-brutality case in a near-future where smartphone manufacturers have begun building video-authentication technology into only some high-end models. The only eyewitness is a passerby who stopped and recorded video and audio of the incident on a cellphone. The video and audio from the defendant officers' body-worn cameras, of a model that authenticates video at capture,[159] seems to cast doubt on the plaintiff's story despite being shaky, blurry, and muffled. Meanwhile, the bystander's cellphone video is much clearer and appears to confirm the plaintiff's version of events. But the cellphone is a cheap, basic model that does not authenticate video at capture.[160] While the officers' body-worn camera footage is readily admitted into evidence, their counsel accuses the witness's cellphone footage of being a high-quality deepfake. Both sides testify, but it's the officers' word against the plaintiffs. The plaintiff cannot afford to pay expert witness fees, and no expert is available to work on the case pro bono. Additionally, the plaintiff's attorney concludes that the bystander-witness would be susceptible to impeachment if called to the stand. After consulting with the client, the plaintiff's attorney makes the difficult decision to withdraw the video.

Any extra up-front cost of video-authentication technology, coupled with the extra litigation cost of proving that a video recorded *without* such technology is *not* fake, could make the difference between justice served and justice denied. Any elevation of the currently low bar for video authentication can be expected to have a disproportionate impact on litigants and witnesses from disadvantaged

made without a reasonable and competent inquiry." Townsend v. Holman Consulting Corp., 914 F.2d 1136, 1140 (9th Cir. 1990). This leaves ample wiggle room when it comes to disputes over evidence.

[156] Tashea, *supra* note 42.

[157] This strategy has an ethical aspect, as discussed in Section VI *infra*.

[158] *Ticks*, *supra* note 25, at 29.

[159] *Id.* at 16 ("For instance, one of the tools in this [video-authentication] industry, Amber Authenticate, works mainly with law enforcement in the United States to integrate their technology within the body cams of police officers. The footage captured by these officers gathers additional signals of trust and hashes the footage directly onto the blockchain.").

[160] *See id.* at 11 ("In a nutshell, with controlled capture, an image, video or audio recording is cryptographically signed, geotagged, and timestamped [at the moment of recording]").

socioeconomic backgrounds while favoring wealthier individuals and corporations (as they have deeper pockets to adopt cutting-edge technologies and weather increased litigation costs), thereby replicating and exacerbating existing disparities in access to justice.[161]

Hearteningly, most existing video-authentication tools are free-of-charge,[162] which lowers the barrier to adoption. Paradoxically, however, the more popular such tools become, the more grounds there will be to cast doubt on any videos *not* recorded with them.[163] If the deepfakes phenomenon leads to heightened expectations to prove authenticity and credibility, genuine but low-tech video evidence will be a casualty of that move away from the current "seeing is believing" paradigm.[164]

D. Deepfakes and Professional Responsibility

Deepfakes pose ethical pitfalls for litigators, including both those attempting to introduce a video into evidence and those seeking to exclude what the other side offers.

The proponent of video evidence must be vigilant against forgeries. As deepfakes become more prevalent, it will be more important than ever for attorneys to verify a video's authenticity as early as possible. If a "smoking gun" video seems too good to be true, it probably is. Importantly, an attorney may not knowingly offer a deepfake into evidence,[165] and may refuse to offer a video

[161] The prohibitively high costs of civil litigation in America have "caused an access to justice problem—people with potentially meritorious claims lack the 'key to the courthouse door'" Simply because they cannot afford to fully litigate the issue. Sasha Nichols, *Access to Cash, Access to Court: Unlocking the Courtroom Doors with Third-Party Litigation Finance*, 5 U.C. IRVINE L. REV. 197, 198 (2015) (footnotes omitted) (listing costs that include "court fees, lawyers' fees, bond requirements, and expert witness fees"). The "basic market inequality" in the provision of legal services means that "individuals, despite suffering a legal harm, are blocked from legal redress because they are too poor to pay for a lawyer." Scott L. Cummings, *The Pursuit of Legal Rights—and Beyond*, 59 UCLA L. REV. 506, 523 (2012). As a result, "[m]illions of Americans lack any access to justice." Deborah Rhode, *Access To Justice: A Roadmap For Reform*, 41 FORDHAM URB. L.J. 1227, 1228 (2014) (footnote omitted) ("Over four-fifths of the poor's legal needs and two- to three-fifths of the legal needs of middle-income Americans remain unmet.").

[162] *Ticks*, *supra* note 25, at 14 (most of the authentication tools surveyed for the report are either free "or offer a free version").

[163] *Ticks*, *supra* note 25, at 28 ("[T]hose who cannot or choose not to use [video authentication] technology"—or who simply did not happen to do so when hurriedly recording video in the heat of the moment—"might find themselves at a disadvantage entering the courtroom, as their credibility may be questioned.").

[164] *Id.*; Maras & Alexandrou, *supra* note 42, at 257 (currently, "[p]eople tend to believe what they see.") (citation omitted).

[165] MODEL RULES OF PROF'L CONDUCT r. 3.3(a)(3) (AM. BAR ASS'N 2019) ("A lawyer shall not knowingly . . . offer evidence that the lawyer knows to be false."); Nix v. Whiteside, 475 U.S. 157, 166 (1986) ("Although counsel must take all reasonable lawful means to attain the

she reasonably believes is a deepfake.[166] If a client pushes an attorney to go forward with a video that attorney suspects or knows is a deepfake, the attorney should consult (as applicable) her firm's ethics counsel, her state bar association's ethics hotline, her state attorney general, or some other authoritative resource concerning professional responsibility.[167]

The duty not to offer a known deepfake as evidence is pretty straightforward.[168] The ethical issues are a bit more nuanced when it comes to challenging the opposing party's evidence. Rooting out and excluding fake evidence preserves the integrity of the judicial process. But when doubt about the authenticity of real evidence starts to pervade the minds of juries, that attitude risks undermining public trust in the courts' truth-finding function. This is why attorneys must tread very carefully when weighing whether to accuse the other side's evidence of deepfakery.

Attorneys should not impugn the authenticity of a video that has been duly authenticated and admitted into evidence, where the attorney does not reasonably believe it is a fake and simply wants to weaken the other side's case in the eyes of the jury. Indeed, to do so would likely cross an ethical line: attorneys are not supposed to make frivolous arguments,[169] make baseless denials of factual contentions,[170] or engage in motion practice simply "to harass, cause unnecessary delay, or needlessly increase the cost of litigation."[171]

Yes, attorneys have an ethical duty to zealously represent their clients.[172] But the bar also expects attorneys to uphold the sanctity of the judicial system as a whole.[173] Lawyers should look beyond the near-term goal of victory in a particular case, or even just a particular motion hearing, and consider the larger impact that cynical deepfake accusations could have on our legal system.

When photographs lie, Susan Sontag wrote in the 1970s, there must be stiff consequences, because photographs "make a claim to be true": "[a] fake photograph (one which has been retouched or tampered with, or whose caption

objectives of the client, counsel is precluded from taking steps or in any way assisting the client in presenting false evidence or otherwise violating the law.").

[166] MODEL RULES OF PROF'L CONDUCT r. 3.3(a)(3) (AM. BAR ASS'N 2019) ("A lawyer may refuse to offer evidence . . . that the lawyer reasonably believes is false.").

[167] *See id.* r. 1.6(b)(4) (permitting attorneys to reveal information relating to a client representation to the extent necessary to ensure compliance with the lawyer's other ethical responsibilities).

[168] *Id.* r. 3.3(a)(3) ("A lawyer shall not knowingly . . . offer evidence that the lawyer knows to be false.").

[169] *Id.* r. 3.1; FED. R. CIV. P. 11(b)(2).

[170] FED. R. CIV. P. 11(b)(4).

[171] *Id.* 11(b)(1).

[172] MODEL RULES OF PROF'L CONDUCT preamble ¶ 2 (AM. BAR ASS'N 2019).

[173] *Id.* ¶ 6 ("a lawyer should further the public's understanding of and confidence in the rule of law and the justice system because legal institutions in a constitutional democracy depend on popular participation and support to maintain their authority").

is false) falsifies reality."[174] In keeping with this responsibility, "news outlets and other publishers of photographs have gone on to establish policies and make decisions regarding the images they use with an eye toward fostering their audience's trust."[175]

Fostering trust is no less important to the judicial system than to the Fourth Estate. That important task falls to judges and lawyers as the principal players in that system. This requires both judges and lawyers to urgently protect the vital public trust that courts can find the truth, and indeed that the truth even exists at all.

This means walking a fine line when it comes to authenticating video evidence at a time when many are also hypervigilant over the perceived threat of deepfakes. It is desirable for juries, and the public in general, to be media-literate. In an age of technological trickery, society stands to benefit if people no longer reflexively believe everything they see and instead learn "to use other indicators — such as trustworthiness of the source — to make informed decisions about whether an image presented is authentic."[176] But if healthy skepticism crosses the line into defeatist nihilism, everyone in the courtroom loses.

Conclusion

Deepfakes will soon make trial attorneys' and judges' jobs more difficult. They will complicate normal trial proceedings, and may give courts reason to revisit the continued adequacy of current rules and standards governing digital evidence. Lawyers will have to exercise greater diligence in verifying the authenticity of video evidence. That includes learning the signs of a deepfake (which will change over time as the technology evolves), consulting a forensic expert when needed, and managing the client friction these measures may cause. Proper diligence before offering a video will shake out fake "chaff" and help the "wheat" survive any authentication challenge.

While I believe there is no need to alter the current rules for authenticating digital video evidence, vigilance against deepfakes will come at a cost. Authenticating videos against deepfake suspicions will prolong litigation and run up costs through extra diligence, additional motion practice and time in court (thereby delaying or extending trial), increased expenditures on lay and expert witnesses, and all the extra billable hours associated with those measures.

There is some risk that deepfakes will pose an existential threat to our democratic institutions, as Professors Citron and Chesney predicted. As deepfake technology improves and it becomes harder to tell what's real, video

[174] Sontag, *supra* note 1, at 66.

[175] Westling, *supra* note 47 (going on to describe the 2003 example of the *Los Angeles Times* "quickly fir[ing] a reporter who had digitally altered Iraq War photographs because the editors realized that publishing a manipulated image would diminish their reader's [sic] perception of the paper's veracity.") (citations omitted).

[176] *Id.*

evidence may lose its customary credibility to juries. If skepticism becomes sufficiently pervasive, it could have a corrosive effect on the justice system as a whole. After all, the courts' core objectives are truth and justice. If juries cease believing that the truth exists and that it can be found out, then they will have little cause to keep believing in the courts. Nor will litigants who are denied the justice they seek because the evidence they presented, while authentic, did not pass technological muster.

That said, rumors of the death of truth are still greatly exaggerated. There is reason for optimism. The courts have long maintained their immunity against infection by previous generations of false evidence. They will survive deepfakes too. With thoughtful advance preparation, trial lawyers and judges will be equipped to handle this new challenge.