Vanderbilt Journal of Entertainment & Technology Law

Volume 23
Issue 2 *Issue 2*

Article 5

2-2021

The Threat of Deepfakes in Litigation: Raising the Authentication Bar to Combat Falsehood

Agnieszka McPeak

The Threat of Deepfakes in Litigation: Raising the Authentication Bar to Combat Falsehood

Agnieszka McPeak[*]

ABSTRACT

Deepfakes are all over the internet—from shape-shifting comedians and incoherent politicians to disturbingly realistic fake pornography. Emerging technology makes it easier than ever to create a convincing deepfake. What used to take significant time and money to develop is now widely available, often for free, thanks to rapid advances in deepfake technology.

Deepfakes threaten individual rights and even democracy. But their impact on litigation should not be overlooked. The US adversarial system of justice is built on a foundation of seeking out the truth to arrive at a just result. The Federal Rules of Evidence serve as an important framework for this truth-seeking mission, and the authentication rules, in particular, should play a key role in preventing deepfake evidence from corrupting the legal process.

This Article looks at the unique threat of deepfakes and how the authentication rules under the Federal Rules of Evidence can adapt to help deal with these new challenges. It examines authentication standards that have emerged for social media evidence and suggests a middle-ground approach that redefines the quantity and quality of circumstantial evidence necessary for a reasonable jury to determine authenticity in the age of deepfakes. This middle-ground approach may raise the evidentiary bar in some cases, but it seeks to balance efficiency with the need to combat falsehood in the litigation process.

[*] Agnieszka McPeak, Associate Professor, Associate Dean for Faculty Scholarship, and Director of the Center for Law, Ethics, and Commerce at Gonzaga University School of Law. Thank you to the editors of the *Vanderbilt Journal of Entertainment and Technology Law* for their expertise, careful edits, and compassion.

TABLE OF CONTENTS

I. INTRODUCTION

The US adversarial system of justice is built on a foundation of seeking out the truth to arrive at a just result. In both civil and criminal cases, the Federal Rules of Evidence serve as a final gatekeeper to funnel potential evidence to that which is relevant, authentic, and not unfairly prejudicial.[1] The Federal Rules of Evidence have always had to grapple with keeping out false information, from forgeries to photoshop.[2] Deepfakes, however, are emerging as a new, sophisticated form of realistic-seeming fabrications. Deepfake technology is rapidly improving in accuracy, speed, and volume. Doctored images that formerly required significant time and money to create can now be made in moments on a smartphone app by almost anyone.[3] As deepfake technology rapidly advances, the naked eye will eventually be unable to discern the subtle clues that indicate an image, video, or audio clip is fake. Even experts struggle with ascertaining the veracity of a potential deepfake.[4] While videos and audio remain powerful pieces of evidence, their authenticity will be hard to gauge in the era of deepfakes.

---

1. *See, e.g.*, FED. R. EVID. 101; FED. R. EVID. 102; FED. R. EVID. 401; FED. R. EVID. 403; FED. R. EVID. 901.

2. *See* Zachariah B. Parry, *Digital Manipulation and Photographic Evidence: Defrauding the Courts One Thousand Words at a Time*, 2009 U. ILL. J.L. TECH. & POL'Y 175, 178–79 (2009).

3. *See* Robert Chesney & Danielle K. Citron, *Disinformation on Steroids*, COUNCIL ON FOREIGN RELS. (Oct. 16, 2018), https://www.cfr.org/report/deep-fake-disinformation-steroids [https://perma.cc/D8U6-NVKX]; *see, e.g.*, Anya Zhukova, *7 Best Deepfake Apps and Websites*, ONLINE TECH TIPS (Aug. 24, 2020), https://www.online-tech-tips.com/cool-websites/7-best-deep-fake-apps-and-websites/ [https://perma.cc/52WZ-RZ9M].

4. *See* Chesney & Citron, *supra* note 3; *see also* Drew Harwell, *Top AI Researchers Race to Detect 'Deepfake' Videos: 'We Are Outgunned*,*'* WASH. POST (June 12, 2019, 4:44 PM), https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/ [https://perma.cc/2V2V-SVMK].

This Article explores the potential impact of deepfakes on the litigation process. It first discusses why deepfakes pose a new and unique threat in litigation and analyzes how the Federal Rules of Evidence attempt to weed out fake content, particularly as to the authentication of social media evidence. It concludes that the Federal Rules of Evidence are generally equipped to handle deepfakes as a new frontier of false evidence, but a higher bar for authentication should be considered in many cases. However, an across-the-board high evidentiary bar may not be warranted, especially in light of cost and efficiency concerns. This Article thus recommends that a middle-ground approach should be used to provide for both flexibility and adequate information to allow a reasonable jury to better gauge the authenticity of potential deepfake evidence.

II. THE UNIQUE THREAT OF DEEPFAKES

With its seemingly endless capacity for creativity, the internet has produced some hilariously realistic celebrity deepfakes. Comedian Bill Hader shape-shifts to take on the faces of the famous actors he is mimicking during a David Letterman interview.[5] Former President Barack Obama gives an obscenity-laced speech about the threat of misinformation to democracy and ends with the plea to "stay woke, bitches."[6] Viral videos of cute babies are doctored to show Elon Musk's face superimposed on the baby's face, leading to predictably creepy results.[7] A search for Nicholas Cage on Reddit produces countless videos of his face added to clips of famous films in which he never starred.[8]

---

5. Jon Blistein, *Watch Bill Hader Become Tom Cruise, Seth Rogen in Eerie Deepfake Video*, ROLLING STONE (Aug. 13, 2019, 4:03 PM), https://www.rollingstone.com/culture/culture-news/bill-hader-tom-cruise-seth-rogen-deepfake-871154/ [https://perma.cc/5L99-PXD8].

6. Todd Spangler, *Jordan Peele Teams with BuzzFeed for Obama Fake-News Awareness Video (Watch)*, VARIETY (Apr. 17, 2018, 11:45 AM), https://variety.com/2018/digital/news/jordan-peele-obama-fake-news-video-buzzfeed-1202755517/ [https://perma.cc/S8G8-WRR4]. The video was created by BuzzFeed and comedian Jordan Peele, who did the voice impersonation used in the video. *Id.* University of Washington researchers have developed an AI tool that allows them to easily manipulate a video of Barack Obama to swap out the speech he is giving, producing a realistic deepfake video. Adam Mann, *Deepfake AI: Our Dystopian Present*, LIVE SCI. (Sept. 30, 2019), https://www.livescience.com/deepfake-ai.html [https://perma.cc/HU3C-XP7B].

7. *See* Amanda Kooser, *This Elon Musk Deepfake Baby Video Shattered My Brain*, CNET (May 10, 2019, 12:23 PM), https://www.cnet.com/news/this-elon-musk-deepfake-baby-video-shattered-my-brain/ [https://perma.cc/QP33-PWM7].

8. *See* Sam Haysom, *People Are Using Face-Swapping Tech to Add Nicholas Cage to Random Movies and What Is 2018*, MASHABLE (Jan. 31, 2018), https://mashable.com/2018/01/31/nicolas-cage-face-swapping-deepfakes/ [https://perma.cc/KJH6-FDQM].

The more disturbing reality, however, is that the vast majority of deepfakes are pornography.[9] Female celebrity faces are being digitally added to pornographic content, creating deepfake porn videos.[10] Kristen Bell,[11] Scarlett Johansson,[12] and Taylor Swift[13] have all been victims of deepfake pornography. But the deepfake pornography phenomenon is not limited to celebrities. Australian law graduate Noelle Martin discovered that her public social media images were used to create explicit photos and videos of her.[14] In addition to the emotional trauma they endure, victims of deepfake pornography suffer stigmatization, reputational harm, harassment, and even blackmail.[15] Women who are victims of abusive deepfake pornography already know the serious impact deepfake technology can have on individual lives.

The threat of deepfakes is amplified by their accuracy, ease of creation, and impact on viewers. Deepfakes are created using artificial intelligence and deep-learning technology.[16] These deepfake applications use real images, audio, and video to generate

---

9. Cleo Abram, *The Most Urgent Threat of Deepfakes Isn't Politics. It's Porn.*, VOX (June 8, 2020, 12:10 PM), https://www.vox.com/2020/6/8/21284005/urgent-threat-deepfakes-politics-porn-kristen-bell [https://perma.cc/3855-N8W5].

10. *See id.*

11. Claudia Willen, *Kristen Bell Says She Was 'Shocked' to Learn That Her Face Was Used in a Pornographic Deepfake Video*, INSIDER (June 11, 2020, 12:16 PM), https://www.insider.com/kristen-bell-face-pornographic-deepfake-video-response-2020-6 [https://perma.cc/7FW4-Y5B6].

12. Isobel Asher Hamilton, *Scarlett Johansson Says Trying to Stop People Making Deepfake Porn Videos of Her Is a 'Lost Cause,'* BUS. INSIDER (Dec. 31, 2018, 4:51 AM), https://www.businessinsider.com/scarlett-johansson-stopping-deepfake-porn-of-me-is-a-lost-cause-2018-12 [https://perma.cc/HE4M-S92V].

13. Ian Morris, *Deepfake Porn Banned by Reddit and Pornhub After Taylor Swift and Meghan Markle Clips Emerge Online*, FORBES (Feb. 7, 2018, 4:42 PM), https://www.forbes.com/sites/ianmorris/2018/02/07/deepfake-porn-banned-by-reddit-and-pornhub-after-taylor-swift-and-meghan-markle-clips-emerge-online/#5a32524a48ea [https://perma.cc/APX7-7CR3].

14. *See* Daniella Scott, *Deepfake Porn Nearly Ruined My Life*, ELLE (June 2, 2020), https://www.elle.com/uk/life-and-culture/a30748079/deepfake-porn/ [https://perma.cc/4D27-2JDL].

15. *See* Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1891–92, 1915, 1924–28 (2019) (explaining the harms of deepfake pornography, including how victims can be manipulated by perpetrators to perform certain acts or pay money to combat widespread dissemination of deepfakes).

16. Kashyap Vyas, *Generative Adversarial Networks: The Tech Behind Deepfake and FaceApp*, INTERESTING ENG'G (Aug. 12, 2019), https://interestingengineering.com/generative-adversarial-networks-the-tech-behind-deepfake-and-faceapp [https://perma.cc/CAC5-AJ7T]. In particular, Generative Adversarial Networks, or GANs, are a form of deep learning that creates, or generates, unique images using data inputs. Outputs then check themselves against the reference data set to help improve realism and accuracy of the generated images. *Id.*

realistic-looking fakes.[17] While movie studios have used sophisticated technology for special effects in film for years, the proliferation of deepfake apps, which provide easy access to sophisticated editing tools, makes deepfakes a new and troubling development.[18] Face-swapping apps and fake video apps, like FakeApp, which allow an average user to create a realistic fake video in minutes for little or no cost, make the democratization of deepfakes possible.[19] This technology is advancing quickly, prompting the quality of deepfakes to continue to increase.[20] Fewer people will be able to detect the visual or aural clues of deepfakes.[21] Deepfake creators will need increasingly smaller data sets to create convincing deepfakes as deep-learning technology advances. At the same time, the internet spurs the swift and broad dissemination of this technology, and more people will have access to the tools for creating deepfakes.[22]

The harmful impact of deepfakes also permeates politics. Millions of people viewed a video of House Speaker Nancy Pelosi appearing to slur her speech while speaking at a press conference.[23] The video was circulated widely on social media, including through a tweet by President Donald Trump.[24] In Malaysia, a rising-star politician was mired in a sex tape scandal after a video surfaced purporting to show him engaging in illegal homosexual activity.[25] Authorities and experts, using facial recognition technology and other forensics, could not establish who actually appears in the video, and all charges were

17. *See* Ben Dickson, *What Is a Deepfake?*, PCMAG (Mar. 4, 2020), https://www.pcmag.com/news/what-is-a-deepfake [https://perma.cc/SF9M-9HTA].

18. *See id.*

19. *See* Kevin Reilly & Steve Kovach, *Face-Swapping Videos Could Lead to More 'Fake News,'* BUS. INSIDER (Apr. 13, 2018, 2:14 PM), https://www.businessinsider.com/fakeapp-lets-people-make-fake-videos-deepfakes-2018-4 [https://perma.cc/7JW4-KW2G].

20. Sharon D. Nelson & John W. Simek, *Detecting Deepfakes*, TECHSHOW, https://www.techshow.com/2020/01/detecting-deepfakes/ [https://perma.cc/JC5C-D9S5] (last visited Dec. 7, 2020).

21. *See id.*

22. *See* Reilly & Kovach, *supra* note 19.

23. Jason Abbruzzese, *Doctored Pelosi Videos Offer a Warning: The Internet Isn't Ready for 2020*, NBC NEWS (May 24, 2019, 1:56 PM), https://www.nbcnews.com/tech/tech-news/doctored-pelosi-videos-offer-warning-internet-isn-t-ready-2020-n1010011 [https://perma.cc/WU4Y-HYNV].

24. Russell Berman, *For Nancy Pelosi, This Is All Just Déjà Vu*, ATLANTIC (May 24, 2019), https://www.theatlantic.com/politics/archive/2019/05/trump-pelosi-video/590233/ [https://perma.cc/C2LR-KES9]. That video was a "shallow fake" that did not rely on deep-learning technology; instead, it involved slowing down the speed of the original and manipulating the audio so that the pitch of the voice remains realistic. *See* Abbruzzese, *supra* note 23.

25. Jarni Blakkarly, *A Gay Sex Tape Is Threatening to End the Political Careers of Two Men in Malaysia*, SBS NEWS, https://www.sbs.com.au/news/a-gay-sex-tape-is-threatening-to-end-the-political-careers-of-two-men-in-malaysia [https://perma.cc/KN2J-ZAX2] (last updated June 17, 2019).

dropped.[26] Nonetheless, the video scandal caused significant political fallout.[27] As a whole, experts have warned that, if unchecked, deepfakes could undermine democracy by amplifying falsehoods and sowing discord.[28]

Deepfakes have also surfaced in litigation,[29] and lawyers will need to take their role in combatting deepfakes seriously as the law tries to deal with this emerging issue.[30] Deepfake evidence has already turned up as a key issue in some cases.[31] For example, a mother in a British court sought to use her husband's threatening audio comments against him in a child custody dispute.[32] Using metadata analysis, the husband was able to show the audio file was a fake, created using software that falsified his voice.[33] While the fake audio file was detected in that case, it serves as a cautionary tale of the power of deepfakes as a source of false evidence.[34]

In the litigation context, two aspects of deepfakes pose a challenge. First, deepfakes may be convincing, compelling, and difficult to detect. Second, people may doubt unaltered content simply because they know realistic deepfakes are possible. First, while the

---

26. *Malaysia's Attorney-General Drops Sex-Video Case; Minister Denounces Plot*, REUTERS (Jan. 9, 2020, 4:33 AM), https://fr.reuters.com/article/us-malaysia-politics-idUSKBN1Z817V [https://perma.cc/PF4X-2DW3]; *see also* A. Ananthalakshmi, *Malaysian Police Say Political Leader Behind Gay Sex Tape Allegations*, REUTERS (July 17, 2019, 1:54 AM), https://www.reuters.com/article/us-malaysia-politics/malaysian-police-say-political-leader-behind-gay-sex-tape-allegations-idUSKCN1UD0OF [https://perma.cc/3RW6-G9GB] (noting that the video appears to be authentic, but facial recognition could not confirm the identity of all parties).

27. *See* Ananthalakshmi, *supra* note 26; *see also* Blakkarly, *supra* note 25.

28. *See* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1769, 1777 (2019) (noting how deepfakes have beneficial applications in education, art, and autonomy); Chesney & Citron, *supra* note 3 (describing the foreign policy implications of deepfakes).

29. *See* Matt Reynolds, *The Judicial System Needs to Learn How to Combat the Threat of 'Deepfake' Evidence*, AM. BAR ASS'N J. (Feb. 28, 2020, 5:11 PM), https://www.abajournal.com/news/article/aba-techshow-experts-warn-of-deepfake-threats-to-justice-system [https://perma.cc/HQ8Z-44SJ].

30. *See* Jason Tashea, *As Deepfakes Make It Harder to Discern Truth, Lawyers Can Be Gatekeepers*, AM. BAR ASS'N J. (Feb. 26, 2019, 7:30 AM), https://www.abajournal.com/lawscribbler/article/as-deepfakes-make-it-harder-to-discern-truth-lawyers-can-be-gatekeepers [https://perma.cc/54CY-NNXP].

31. *See* Matt Reynolds, *Courts and Lawyers Struggle with Growing Prevalence of Deepfakes*, AM. BAR ASS'N J. (June 9, 2020, 10:29 AM), https://www.abajournal.com/web/article/courts-and-lawyers-struggle-with-growing-prevalence-of-deepfakes [https://perma.cc/7ALF-PKVW].

32. *Id.*

33. *Id.*

34. *See id.*

democratization of deepfakes has its upsides,[35] it also brings about a new and troubling era of disinformation. In the free speech context, the general approach has been to assume that the cure for false speech is more speech.[36] But the flood of information online—real, fake, or in between—has shown that individuals are not necessarily able to access counterspeech or otherwise meaningfully seek out the truth.[37] Similarly, in the context of jurors, deepfakes may soon go beyond a layperson's ability to visually ascertain whether an image or video purports to be what it seems.[38] Second, the proliferation of disinformation makes people question their ability to trust anything. Social media disinformation has already made it harder for people to distinguish truth from fiction online.[39] Thus, in addition to the risk of deepfakes being perceived as real, the knowledge that deepfakes are out there undermines belief in the authenticity of undoctored images. In other words, people no longer believe *anything* is real.[40]

Many people value visual perception above other indicators of truth.[41] Images are powerful, and social media has further elevated videos and images as sources of factual information. More Americans get their news from social media than print media.[42] With Instagram, TikTok, and other video- and photo-heavy platforms gaining market share, social media trends continue to elevate visual content—and

---

35. *See* Jessica Silbey & Woodrow Hartzog, *The Upside of Deep Fakes*, 78 MD. L. REV. 960, 960 (2019). Notably, the democratization of deepfakes has its upsides, like education, accessibility, and freedom of expression. *Id*. at 960–61; *see also* Chesney & Citron, *supra* note 28, at 1769 (noting how deepfakes have beneficial applications in education, art, and autonomy).

36. *See* Lyrissa Barnett Lidsky, *Nobody's Fools: The Rational Audience as First Amendment Ideal*, 2010 U. ILL. L. REV. 799, 822 (2010) (noting how the rational audience and concept of counterspeech remain important in First-Amendment jurisprudence).

37. *See generally* Julie A. Seaman, *Black Boxes*, 58 EMORY L.J. 427, 461 (2008) (noting how the US system clings to the jury's role as judge of credibility despite technological advances); John Villasenor, *Artificial Intelligence, Deepfakes, and the Uncertain Future of Truth*, BROOKINGS (Feb. 14, 2019), https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/ [https://perma.cc/WC89-W73R].

38. *See* Villasenor, *supra* note 37 (noting the need for public awareness about deepfakes).

39. *See* Janna Anderson & Lee Rainie, *The Future of Truth and Misinformation Online*, PEW RSCH. CTR. (Oct. 19, 2017), https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/ [https://perma.cc/VH2F-M23Q].

40. *See* Riana Pfefferkorn, *'Deepfakes': A New Challenge for Trial Courts*, NWSIDEBAR (Mar. 13, 2019), https://nwsidebar.wsba.org/2019/03/13/deepfakes-a-new-challenge-for-trial-courts/ [https://perma.cc/A8WL-N2WV].

41. *See* Carolyn Purnell, *Do We All Still Agree that "Seeing Is Believing"?*, PSYCH. TODAY (June 23, 2020), https://www.psychologytoday.com/us/blog/making-sense/202006/do-we-all-still-agree-seeing-is-believing [https://perma.cc/GK2H-B7Q7].

42. Elisa Shearer, *Social Media Outpaces Print Newspapers in the U.S. as a News Source*, PEW RSCH. CTR. (Dec. 10, 2018), https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/ [https://perma.cc/5B94-6B46].

videos in particular—over other formats.[43] But social media also has become fertile ground for the spread of deepfakes.[44] And, as the public becomes more aware of the risk of being fooled by realistic deepfakes, they also begin to mistrust authentic videos too. Take, for example, a video of Gabon's president Ali Bongo. Bongo, who suffered a stroke and was out of the public eye for months, was rumored to be very ill or dead.[45] The video, meant to quash public fears, was attacked as a deepfake that further confirmed speculation about his poor health.[46] Controversy over the video even led to an unsuccessful military coup.[47] To date, speculation continues about whether the video was a deepfake, but the uncertainty and mistrust it spawned serve as a cautionary tale about the public's inability to gauge authenticity in the age of deepfakes.[48] New technology allows deepfakes to become even trickier to detect. Seeing is no longer believing, and viewers may now question even real content.[49]

III. DEEPFAKES & THE FEDERAL RULES OF EVIDENCE

In litigation, deepfakes threaten to undermine the court's fact-finding mission. First, if a deepfake is admitted as authentic evidence, it can undermine the court's truth-seeking mission. Deepfakes can easily become outcome determinative, given that video evidence in the courtroom is impactful and can profoundly influence individual perception of events.[50] Second, deepfakes also undermine the public's ability to trust authentic videos. Judges and juries may be skeptical of believing what they see in a real, undoctored video because they know images are now easily manipulated. In other words, the knowledge that deepfakes are out there and hard to spot may make fact-finders question whether they can even believe real footage.

Fortunately, proper use of the authentication rules in the Federal Rules of Evidence can alleviate both concerns. The key, as

---

43. *See* Deep Patel, *12 Social Media Trends to Watch in 2020*, ENTREPRENEUR (Dec. 20, 2019), https://www.entrepreneur.com/article/343863 [https://perma.cc/7ZKR-MV89].

44. *See, e.g.*, Ali Breland, *The Bizarre and Terrifying Case of the "Deepfake" Video That Helped Bring an African Nation to the Brink*, MOTHER JONES (Mar. 15, 2019), https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/ [https://perma.cc/9VSS-FD8G].

45. *Id.*

46. *Id.*

47. *Id.*

48. *See* Janosch Delcker, *Welcome to the Age of Uncertainty*, POLITICO (Dec. 17, 2019, 7:50 PM), https://www.politico.eu/article/deepfake-videos-the-future-uncertainty/ [https://perma.cc/LSD5-RXB3].

49. Purnell, *supra* note 41; *see also* Chesney & Citron, *supra* note 28, at 1785.

50. *See* Purnell, *supra* note 41.

this Article will demonstrate, is to employ a sufficient standard of authenticity under Federal Rule of Evidence 901.

A. How the Federal Rules of Evidence Weed Out Fakes

The purpose of the Federal Rules of Evidence is "to administer every proceeding fairly, eliminate unjustifiable expense and delay, and promote the development of evidence law, to the end of ascertaining the truth and securing a just determination."[51] They embody the principle that an adversarial system of justice is the ideal mechanism for reaching the truth through litigation. The proponent of evidence bears the burden to establish relevance and authenticity.[52] Other limitations also apply, such as to hearsay evidence and the balancing of the evidence's probative value with unfair prejudice.[53] But even disputed evidence may be admissible, as opposing parties can present competing evidence, cross-examine witnesses, and otherwise seek out the truth throughout the litigation process. It is then for an impartial judge or jury, considering all the evidence, to decide the truth.

One of the fundamental underpinnings of the US legal system is that a jury of the litigants' peers is best equipped to find the truth based on the evidence presented. Unfortunately, individuals may now have a harder time detecting truth from lies online,[54] and deepfakes further challenge the jury's ability to perform its truth-seeking role. Authentication requirements, thus, must serve as a crucial roadblock to the use of deepfakes in litigation.

1. Authentication of Electronically Stored Information

With the advent of electronically stored information (ESI), courts had to consider whether existing admissibility rules were sufficient to address ESI as a new category of potential evidence. In *Lorraine v. Markel American Insurance Co.*,[55] the US District Court for

51. FED. R. EVID. 102.

52. FED. R. EVID. 901(a) (requiring that the proponent of the evidence show that the evidence is what it purports to be); FED. R. EVID. 401 (requiring that evidence must have the tendency to make some fact that is of consequence to the litigation more or less probable).

53. FED. R. EVID. 403; FED. R. EVID. 801; FED. R. EVID. 802. Hearsay is defined in Federal Rule of Evidence 801, FED. R. EVID. 801, with exceptions spelled out in subsequent rules, FED. R. EVID. 802. Rule 403 contains the balancing test that allows otherwise admissible evidence to be excluded on the basis of unfair prejudice outweighing its probative value. FED. R. EVID. 403.

54. *See* Raina Ducklow & Bud Mortenson, *Why People Are Better at Lying Online than Telling a Lie Face-to-Face*, SCI. DAILY (May 5, 2009), https://www.sciencedaily.com/releases/2009/05/090503203738.htm [https://perma.cc/LP64-VCX6].

55. Lorraine v. Markel Am. Ins. Co., 241 F.R.D. 534 (D. Md. 2007).

the District of Maryland identified the ways in which existing admissibility rules apply to ESI.[56] The *Lorraine* opinion first provides an overview of evidentiary hurdles that parties must cross before ESI can be admitted, including relevance considerations under Rule 401.[57] Additionally, relevant ESI must survive the rules prohibiting hearsay, the original writing rule, and the Rule 403 balancing of probative value against the danger of unfair prejudice.[58]

Even if otherwise admissible under the Rules, ESI evidence still must be authentic. Authentication is outlined in Rule 901, which states that the proponent of evidence "must produce evidence sufficient to support a finding that the item is what the proponent claims it is."[59] Parties may satisfy this requirement by presenting the testimony of a witness with knowledge[60] or by using the distinctive characteristics of the evidence, such as "the appearance, contents, substance, internal patterns, or other distinctive characteristics of the item, taken together with all the circumstances."[61] Authentication can also be established with evidence of a process or system, by showing that it "produces an accurate result."[62] For authentication of voice audio, Rule 901(b)(5) requires "[a]n opinion identifying a person's voice—whether heard firsthand or through mechanical or electronic transmission or

---

56. *Id.* at 541; *see also* Paul W. Grimm, Michael V. Ziccardi & Alexander W. Major, *Back to the Future:* Lorraine v. Markel American Insurance Co. *and New Findings on the Admissibility of Electronically Stored Information*, 42 AKRON L. REV. 357, 362–63 (2009).

57. *Lorraine*, 242 F.R.D. at 538 (describing the first two admissibility steps as "(1) is the ESI relevant as determined by Rule 401 (does it have any tendency to make some fact that is of consequence to the litigation more or less probable than it otherwise would be); (2) if relevant under 401, is it authentic as required by Rule 901(a) (can the proponent show that the ESI is what it purports to be)").

58. *Id.* (describing the remaining admissibility steps as "(3) if the ESI is offered for its substantive truth, is it hearsay as defined by Rule 801, and if so, is it covered by an applicable exception (Rules 803, 804 and 807); (4) is the form of the ESI that is being offered as evidence an original or duplicate under the original writing rule, or if not, is there admissible secondary evidence to prove the content of the ESI (Rules 1001-1008); and (5) is the probative value of the ESI substantially outweighed by the danger of unfair prejudice or one of the other factors identified by Rule 403, such that it should be excluded despite its relevance.").

59. FED. R. EVID. 901(a). Notably, Rule 104 also addresses preliminary matters more broadly and allows admission of evidence to occur in two ways. FED. R. EVID. 104. First, the court can make the preliminary decision as to whether evidence is admissible. FED. R. EVID. 104(a). Under this path, the court "is not bound by evidence rules, except those on privilege." *Id.* The second path derives from Rule 104(b): "Relevance that Depends on a Fact." FED. R. EVID. 104(b). Under this path, the fact-finder must be presented with evidence to support that a particular fact exists, and that evidence must be admissible. *Id.* Of these two paths, it is Rule 104(b)'s "relevance that depends on a fact" that often applies to authenticity determinations. *See* Grimm et al., *supra* note 56, at 364.

60. FED. R. EVID. 901(b)(1).

61. FED. R. EVID. 901(b)(4).

62. FED. R. EVID. 901(b)(9).

recording—based on hearing the voice at any time under circumstances that connect it with the alleged speaker."[63] Proper authentication of digital videos or photographs may require detailed evidence about chain of custody, such as how illegal pornographic content was retrieved from a defendant's computer and subsequently stored.[64] Testimony of someone who accessed content from the internet is insufficient to attribute content to a particular user, without "personal knowledge of who maintains the website, who authored the documents, or the accuracy of their contents."[65]

The threshold for making a prima facie showing of authenticity is not high, however, and it suffices to merely offer "a foundation from which a jury could reasonably find that the evidence is what the proponent says it is."[66] This burden is slight.[67] Ultimately, the proponent of digital evidence can usually authenticate that evidence with other admissible evidence supporting its genuineness.[68] This basic approach continues to be used for digital evidence as well.

While courts have been able to apply the existing Federal Rules of Evidence to ESI, the rise of deepfakes marks a new era of altered and fabricated evidence, and a higher bar may be necessary.[69] The technology to create deepfakes will continue to outpace the knowledge and ability of judges, lawyers, and laypeople alike.[70] A greater responsibility will fall on lawyers to challenge evidence that is a deepfake. Further, because the proponent of the evidence bears the burden of establishing authenticity, litigants will have to contend with their own attempts to rely on favorable evidence that may be difficult

---

63. FED. R. EVID. 901(b)(5).

64. *See* United States v. Salcido, 506 F.3d 729, 733 (9th Cir. 2007) (holding that the government properly authenticated child pornography taken off of the defendant's computer by presenting detailed evidence as to the chain of custody, specifically how the images were retrieved from the defendant's computers).

65. Wady v. Provident Life & Accident Ins. Co. of Am., 216 F. Supp. 2d 1060, 1064 (C.D. Cal. 2002).

66. United States v. Safavian, 435 F. Supp. 2d 36, 38 (D.D.C. 2006) (quoting 5 FEDERAL RULES OF EVIDENCE MANUAL § 901.02[1] (2020)) (discussing authentication of emails).

67. *See, e.g.*, United States v. Reilly, 33 F.3d 1396, 1404 (3d Cir. 1994) ("[T]he burden of proof for authentication is slight.").

68. *See* Daniel Capra, *Authenticating Digital Evidence*, 69 BAYLOR L. REV. 1, 3 (2017).

69. *Cf.* Grimm et al., *supra* note 56, at 366. ("As electronic evidence becomes more ubiquitous at trial, it is critical for courts to start demanding that counsel give more in terms of authentication—and counsel who fail to meet courts' expectations will do so at their own peril.").

70. *Cf. id.* ("But it is also a very real possibility that someone *inept* with computers may also alter electronic evidence so as to make it unusable or inadmissible. Therefore, as technology continues creating relevant evidence while, simultaneously, outpacing the working knowledge and ability of most lawyers and judges to deal with it, ensuring proper authentication of electronic evidence becomes a greater responsibility for attorneys and judges alike.").

to prove is real. For example, in 2002, the US Supreme Court in *Ashcroft v. Free Speech Coalition* considered the constitutionality of an anti-child pornography statute that attempted to prohibit images that merely "appear to be" or "convey the impression of" a minor engaging in sexually explicit conduct.[71] When arguing in favor of the statute, the government emphasized the challenges of proving a pornographic image contains a real person and is not a fake.[72] The statute thereby contemplates banning both real and virtual pornography: "As imaging technology improves, Congress found, it becomes more difficult to prove that a particular picture was produced using actual children. To ensure that defendants possessing child pornography using real minors cannot evade prosecution, Congress extended the ban to virtual child pornography."[73] Nonetheless, the Court held that the statute was overbroad and unconstitutional because it attempted to ban material that was neither obscene under the law nor exploiting real children.[74]

While courts have found ways to apply the existing evidence rules to ESI generally, deepfakes pose a new and complicated challenge to the authentication process. The relatively low bar for authentication may be insufficient as deepfakes continue to rise in prominence. Social media content, in particular, has forced more scrutiny on the sufficiency of current authentication rules.

2. Authentication of Social Media Evidence

Because deepfakes are often obtained and shared on social media, a look at the authentication standards for social media evidence is instructive. Over the last decade, courts have taken differing approaches to social media evidence authentication. Some jurisdictions, like Maryland, have created a higher bar for authenticating social media content.[75] Others, like Texas, have treated social media evidence the same as any other content, thus requiring less proof of authenticity.[76]

Under the higher-bar Maryland approach, courts have noted the ease with which false information can be created through fictitious accounts or unauthorized access to real accounts.[77] While

71. *See* Ashcroft v. Free Speech Coal., 535 U.S. 234, 242–43 (2002).

72. *See id.* at 242.

73. *See id.*

74. *See id.* at 256–58.

75. *See* Brendan W. Hogan, Note, Griffin v. State: *Setting the Bar Too High for Authenticating Social Media Evidence*, 71 MD. L. REV. ENDNOTES 61, 61 (2012).

76. *See* Tienda v. State, 358 S.W.3d 633, 638–39 (Tex. Crim. App. 2012).

77. *See* Griffin v. State, 19 A.3d 415, 421 (Md. 2011).

circumstantial evidence about "distinctive characteristics" of the evidence sometimes suffices for authenticity for other types of evidence, social media evidence cannot be authenticated merely by the name and biographical details of the account holder.[78] Rather, courts have enumerated three methods of authenticating social media evidence. First, a witness with requisite knowledge can testify that the social media evidence is what it claims to be.[79] Second, an expert can perform computer forensic searches of the device used to create the content to identify where the content originated.[80] Third, the social media website itself can directly provide some evidence that links a profile to the person or the content to its creator.[81]

Griffin v. State, which established the Maryland approach, involved a MySpace comment that the state attempted to authenticate only through the testimony of the investigator who found it online.[82] However, the state failed to get testimony of the post's author or other witnesses with knowledge of its creation and thus failed to authenticate the evidence.[83] Similar authentication issues have arisen in other jurisdictions. For example, in *United States v. Vayner*, the US Court of Appeals for the Second Circuit vacated and remanded a conviction for unlawful transfer of a falsified identification document after the lower court improperly admitted a social media page printout.[84] A key issue during the trial was whether the defendant could be linked to the Gmail account that sent a forged birth certificate.[85] The government presented a printout of a profile page from a Russian social media site akin to Facebook as its key piece of evidence.[86] The printout contained the defendant's image and a version of his name.[87] It also contained a username for Skype, a video-messaging platform, and the Gmail address that matched the one that sent the forged document.[88] In attempting to authenticate the document, the government presented testimony from the special agent who accessed the profile via the internet.[89] But the government presented no evidence that the

78. *See id.* at 423–24.
79. *See id.* at 427.
80. *See id.* at 427–28.
81. *See id.* at 428.
82. *Id.* at 418.
83. *Id.* at 423–24.
84. United States v. Vayner, 769 F.3d 125, 127 (2d Cir. 2014).
85. *Id.* at 132.
86. *See id.* at 128.
87. *Id.*
88. *Id.* at 128–29.
89. *Id.* at 127–28.

defendant created the profile, that the platform confirms user identities for profiles, or that the Gmail account listed in the profile belonged to the defendant.[90] Therefore, the evidence did not establish that the document was genuine and was too speculative to support "a reasonable conclusion that this page was created by the defendant or on his behalf."[91] The court then held the error was not harmless, and vacated and remanded the case.[92]

Notably, the *Vayner* court did not state what kind of evidence would have sufficed to authenticate the social media printout. Instead, the court merely noted that many ways exist to authenticate documents and no type or quantum of authentication evidence is expressly defined in the Federal Rules of Evidence.[93] The profile page at issue was a key piece of evidence corroborating another witness's story, and the government did not do enough to show that the defendant created the page or that it accurately reflected his Gmail account address.[94]

By contrast, the Maryland approach ultimately creates a higher bar, requiring the proponent of social media evidence to prove its authenticity through more definitive means such as the testimony of the creator, a forensic expert, or the hosting platform.[95] Biographical data on the social media site and similar characteristics are not distinctive enough to trigger the jury's consideration of the factual basis for claimed authenticity.[96]

While the Maryland approach establishes a higher bar, other jurisdictions, like Texas, take a more flexible approach. Texas only requires the party proffering the evidence to make a threshold showing of authenticity.[97] It is then for the fact-finder to ultimately make the determination.[98] For example, in *Tienda v. Texas*, the court held that MySpace posts were properly authenticated because the state offered a combination of circumstantial evidence to help establish that the posts were made by the defendant.[99] These facts included multiple distinctive

90. *Id.* at 131–32.

91. *Id.* at 132.

92. *Id.* at 134–35.

93. *Id.* at 133.

94. *Id.*

95. *See* Griffin v. State, 19 A.3d 415, 427–28 (Md. 2011); *see also* Hogan, *supra* note 75, at 79.

96. *See* Paul W. Grimm, Lisa Yurwit Bergstrom & Melissa M. O'Toole-Loureiro, *Authentication of Social Media Evidence*, 36 AM. J. TRIAL ADVOC. 433, 448 (2013) (describing cases that have excluded social media evidence, or evidence as to its purported authenticity, due to a lack of threshold showing).

97. *See* Tienda v. State, 358 S.W.3d 633, 638–39 (Tex. Crim. App. 2012).

98. *Id.* at 638.

99. *Id.* at 645.

photographs of the defendant, references to specific facts and people, and the listing of the defendant's email address.[100] Based on these facts, the court held that this circumstantial evidence, taken as a whole, "justif[ied] admitting the evidence and submitting the ultimate question of authenticity to the jury."[101]

The Texas approach thus avoids elevating the standard for admissibility of social media evidence above the usual approach under the Federal Rules of Evidence. The authentication of social media evidence is a determination based on whether there is sufficient evidence for a reasonable jury to conclude that the evidence is authentic. But courts have "historically considered admissibility of all documentary evidence on a continuum, in which clearly authentic evidence is admitted, clearly inauthentic evidence is excluded, and everything in between is conditionally relevant and admitted for the jury to determine its authenticity."[102] The Federal Rules of Evidence thus already function in a way that allows jurors to ascertain the authenticity of evidence when it is not clearly authentic or inauthentic. In this way, the higher bar in the Maryland approach has been criticized as too onerous, while the more flexible Texas approach is touted as sufficient even for the new developments in how we create and share electronic content.[103]

B. Finding Middle Ground for Assessing Deepfake Evidence

The flexible standard used by the Texas approach for authenticating digital evidence, and specifically social media evidence, presupposes that jurors will have enough circumstantial evidence to reasonably judge authenticity. With deepfakes, simply identifying the source of a video or audio and leaving jurors to use their senses to gauge genuineness may not suffice, particularly as deepfakes become more sophisticated.[104] Thus, the Texas approach may prove too lenient to deal

100. *Id.*

101. *Id.* at 647; *accord* People v. Valdez, 135 Cal. Rptr. 3d 628, 632–34 (2011).

102. Grimm et al., *supra* note 96, at 456–57 (advocating for the more flexible approach to social media evidence authentication).

103. *See, e.g.*, Hogan, *supra* note 75; John Patzakis, *Delaware Court Affirms Conviction Based on Facebook Evidence*, NEXT GENERATION EDISCOVERY L. & TECH BLOG (Mar. 5, 2014, 9:45 AM), https://blog.x1discovery.com/tag/tienda-v-state/ [https://perma.cc/Q8RW-MD5L] (stating that the authentication test from *Tienda*, the 2012 Texas Court of Criminal Appeals case, has become the majority view in the United States).

104. *See* Elizabeth Caldera, *"Reject the Evidence of Your Eyes and Ears": Deepfakes and the Law of Virtual Replicants*, 50 SETON HALL L. REV. 177, 189 (2019) (explaining some of the subtle ways deepfake content can be used, like lack of blinking by the subject, but noting that the technology will make it even harder to rely on visual clues in the future).

with the new threat of deepfakes. Something more rigorous than the Texas approach is necessary.

The Maryland approach, however, may be too high a bar. For example, it may be inefficient to require analysis of metadata and the use of technology experts in many cases. Deepfake detectors may be costly, requiring computer forensics or other expert analysis. Further, while detecting deepfakes through technology is certainly possible, it is imperfect, especially as deepfake technology continues to advance.[105] A blanket requirement for forensic expert testimony would unnecessarily increase the cost of litigation. The struggle thus becomes finding cost-effective tools for weeding out deepfakes while maintaining rigor in gauging authentication, veracity, and balancing the video image's relevance with its unfair prejudice.

The solution may thus lie in a middle-ground approach that redefines the quantity and quality of circumstantial evidence necessary for a reasonable jury to determine authenticity in the age of deepfakes. Some courts have noted that social media evidence contains an inherent risk of falsification.[106] Now, more than ever, the risk of convincing fakes is on the rise with the democratization of deepfakes. Simply relying on basic human perception no longer suffices with this new genre of falsehoods. Perhaps this is best illustrated with aural evidence. The current version of Rule 901 contemplates lay opinion identifying a voice as the alleged speaker, without taking into account the reality that current deepfake technology can accurately simulate a speaker's voice.[107] Even an individual familiar with another person's voice can be easily fooled by a deepfake audio clip. Thus, it is important that other circumstantial evidence is presented to authenticate some audio clips to allow a reasonable jury to determine admissibility.

Under this middle-ground approach, circumstantial evidence should provide particular context of how a video or image originated,

---

105. *See* Robert Chesney & Danielle Keats Citron, *21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security*, 78 MD. L. REV. 882, 889 (2019) (describing how technologists cannot solve all of our deepfake-related problems); Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J.L. & TECH. 1, 57–58 (2020) (noting that some sort of flagging technology could be developed to detect deepfakes but would be an incomplete fix without greater media literacy, platform social responsibility, and other solutions); Chesney & Citron, *supra* note 28, at 1787–88 (noting how PhotoDNA or similar technology for deepfake detection is underdeveloped and may not be a feasible solution to the prevalence of deepfakes more broadly); Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 353–55 (2019) (describing how the private sector can help counter deepfakes).

106. *See* State v. Eleck, 23 A.3d 818, 822 (Conn. App. Ct. 2011) (noting how easy it is to fake social media content due to lack of security on platforms).

107. *See* FED. R. EVID. 901(b)(5).

who it purports to depict, and what features of the video or image support authenticity—without necessarily requiring computer forensics and analysis in every case. Testimony of those with knowledge may become even more important. But judges will need to assume a strong gatekeeping role to ensure that the quantum of circumstantial evidence is sufficient for a jury to make a more nuanced decision about authenticity in the age of deepfakes.[108] A special jury instruction about visual and audio clues for detecting deepfakes may be helpful in some situations.[109] A middle-ground approach should promote fairness and efficiency without creating too high a bar in most cases. Fortunately, the Federal Rules of Evidence already contemplate balancing competing concerns to arrive at the truth and, with appropriate standards for authenticity, potential deepfakes will be less likely to

---

108. *See generally* FED. R. EVID. 104(a) (stating that the court has the responsibility to decide any preliminary questions regarding the admissibility of evidence).

109. Juries are often instructed on how to gauge the credibility of witnesses. For example, the US Court of Appeals for the Third Circuit uses a model jury instruction that states:

> In deciding what the facts are, you may have to decide what testimony you believe and what testimony you do not believe. You are the sole judges of the credibility of the witnesses. "Credibility" means whether a witness is worthy of belief. You may believe everything a witness says or only part of it or none of it. In deciding what to believe, you may consider a number of factors, including the following:
>
> (1) the opportunity and ability of the witness to see or hear or know the things the witness testifies to;
>
> (2) the quality of the witness's understanding and memory;
>
> (3) the witness's manner while testifying;
>
> (4) whether the witness has an interest in the outcome of the case or any motive, bias or prejudice;
>
> (5) whether the witness is contradicted by anything the witness said or wrote before trial or by other evidence;
>
> (6) how reasonable the witness's testimony is when considered in the light of other evidence that you believe; and
>
> (7) any other factors that bear on believability.
>
> [The weight of the evidence to prove a fact does not necessarily depend on the number of witnesses who testify. What is more important is how believable the witnesses were, and how much weight you think their testimony deserves.]

COMM. ON MODEL CIV. JURY INSTRUCTIONS, THIRD CIR., MODEL CIVIL JURY INSTRUCTIONS § Ch. 1.7 (2018). For videos, images, and audio evidence that needs to be authenticated with circumstantial evidence, the court may opt to include a special jury instruction that explains some criteria that the jurors can use to determine if they believe the evidence purports to be what it claims it is. *See* Agnes E. Venema & Zeno J. Geradts, *Digital Forensics, Deepfakes, and the Legal Process*, 16 SCITECH LAW. 14, 17 (2020). Aspects like lighting, blinking, and editing clues can be included as factors. *See* Caldera, *supra* note 104 (explaining some of the subtle ways deepfake content can be used, like lack of blinking by the subject, but noting that the technology will make it even harder to rely on visual clues in the future). Unfortunately, a jury instruction about ways to spot deepfakes may become obsolete as technology advances. But some sort of detailed guidance for the jury on gauging authenticity may be warranted, at least as a short-term solution.

corrupt the litigation process. This middle-ground approach balances emerging evidentiary needs in the era of deepfakes; it elevates the traditional standard above a basic, flexible approach, which will fail to protect the court against the threat of deepfakes in litigation, while also avoiding an unduly high bar.

IV. CONCLUSION

Deepfake technology has created new challenges in litigation, particularly as to authentication of video and audio evidence. While some courts impose a higher bar for admitting social media evidence, others take a flexible approach that applies the existing Federal Rules of Evidence with little or no modification to account for the ease with which false information is created and shared on social media. Now, with the advent of deepfakes, a more rigorous approach to authenticity will be necessary.

This Article suggests that a middle-ground approach that rests between the current Maryland and Texas standards. This approach would recognize that deepfakes pose unique challenges to the truth-seeking aim of the Federal Rules of Evidence. While an across-the-board heightened standard of authentication for digital evidence is not warranted, a higher bar may be required in some cases. A middle-ground approach would mandate that sufficient circumstantial evidence—both as to quantum and quality—be presented to the jury whenever the jury is asked to ascertain the authenticity of digital video and audio evidence. By requiring this middle-ground approach, the Federal Rules of Evidence can ensure jurors have sufficient evidence to gauge the authenticity of increasingly convincing deepfakes that threaten to undermine the justice system.