MITCHELL | HAMLINE School of Law

Mitchell Hamline Law Review

Volume 48 | Issue 1 Article 5

2022

A New Reality: Deepfake Technology and the World Around Us

Molly Mullen

A NEW REALITY:

DEEPFAKE TECHNOLOGY AND THE WORLD AROUND US

Molly Mullen[1]

I. INTRODUCTION: IS THAT REALLY ME?

"[W]ithout corroboration, there is absolutely no way to know whether somebody's memory is a real memory or a product of suggestion."
-Elizabeth Loftus[2]

You are bickering with your significant other during the height of an argument as your partner begins to hurl corrosive, malicious, and fraudulent accusations your way. You seemingly become more and more confused as your significant other asserts you have done the unthinkable: cheated. As thoughts and concerns swirl around your baffled mind, you offer yourself some comfort as you know it is not true, and hopefully a thorough explanation can clear the air. As you begin defending your chastity, your significant other pulls out their cell phone and shows you something you cannot fathom. There, right there, in physical, audio, and video form, on the six-inch screen of their phone, a video plays of you engaging in salacious acts with another person who is clearly not your significant other. The video is crisp, clear, and undoubtedly looks, acts, and talks like you. But is it *me?*

[1] Thank you to Barry University School of Law and Professor Mitchell Frank for his sponsorship, encouragement, support, and enthusiasm in teaching evidence to law students.
[2] ELIZABETH F. LOFTUS, MEMORY: SURPRISING NEW INSIGHTS INTO HOW WE REMEMBER AND WHY WE FORGET 169 (1980).

You then realize that the confidence you held minutes ago in your own candor is useless. Somehow, someway, there is video proof of you committing an act you are positive you did not do. Was I *drugged?* Did I *blackout* and not remember my actions? Your brain is trying to comprehend and protect itself from the trauma and deceit presented to it, but you realize that this video—this horrifying, false, and life-changing video is real, at least to the beholder. You know that the video bears a resemblance to your likeness, but you also know it is not real.

Nevertheless, the video *is* real, and that is all your significant other needs to know. The damage is done, and now it appears there is irrefutable proof you have committed the abominable act of adultery that you know you did not do. You begin to question your own reality and the world around you. If this false video exists but appears to be genuine in every way, what else in our world is simply a distortion of the truth? What other forms of false media dwell in the world that will swindle and deceive even the most intelligent minds?

This Article explains deepfake technology and reviews deepfakes in modern society.[3] Next, this Article discusses the evidentiary implications of deepfakes and how they are analyzed and authenticated within the courtroom.[4] Finally, this Article discusses potential methods of combatting deepfakes, their effect on the courtroom, and prospective avenues of redress for deepfake victims.[5]

II. WHAT IS A DEEPFAKE?

The disturbing anecdote above is an example of a person procuring and exploiting what is known to be a "deepfake."[6] This example, although life-altering and ultimately distressing, is trivial and uneventful in comparison to the damage that could be done with this technology, which will further be explored after identifying and explaining the technology itself. Deepfakes are videos, images, or other media that have been manipulated to appear as if the subject of the medium is speaking or partaking in an action that they did not actually undertake.[7] Cleverly named, the word deepfake stems from a portmanteau combination of "deep learning" and

---

[3] *See infra* Part II and III.

[4] *See infra* Part IV, V, and VI.

[5] *See infra* Part VII, VIII, IX and X.

[6] *See Words We're Watching: 'Deepfake,'* MERRIAM WEBSTER (Apr. 2020), https://www.merriam-webster.com/words-at-play/deepfake-slang-definition-examples [https://perma.cc/BER4-E8JN]; 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/SC48-Q5MN].

[7] Benjamin Goggin, *From Porn to 'Game of Thrones': How Deepfakes and Realistic-Looking Fake Videos Hit It Big*, BUS. INSIDER (June 23, 2019, 10:45 AM), https://www.businessinsider.com/deepfakes-explained-the-rise-of-fake-realistic-videos-online-2019-6 [https://perma.cc/XY6M-TUZK].

"fake."[8] Deepfakes, which are hyper realistic, are the result of artificial intelligence applications that are able to create a new video or audio clip based on an amalgam of technology. This includes superimposing, altering, and merging images together to create a sophisticated and believable video.[9]

Deepfake technology is relatively new as the technological mechanisms for creating the content are young and developing; but, despite its infancy, deepfakes have already infiltrated the world around us. While currently the most prevalent use of deepfakes to date has been in the pornographic film industry,[10] the world has already seen deepfakes influence the political realm, social media, and education.[11] The ever-growing list of possible implications for the technology is overwhelming.[12]

Deepfakes are synthesized when the video creator takes an image or likeness of one person's face and subsequently replaces it with another face or body using an artificially intelligent facial recognition algorithm.[13] Deepfakes rely on deep learning, which is a "subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neuron networks."[14] Deepfakes are generally synthesized in two ways.[15]

---

[8] Tom Taulli, *Deepfake: What You Need to Know*, FORBES (June 15, 2019, 1:02 PM), https://www.forbes.com/sites/tomtaulli/2019/06/15/deepfake-what-you-need-to-know/ [https://perma.cc/A982-K5S2].

[9] *Id.*; *see also* Goggin, *supra* note 7.

[10] Karen Hao, *Deepfake Porn is Ruining Women's Lives. Now the Law May Finally Ban It.,* MIT TECH. REV. (Feb. 12, 2021), https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/ [https://perma.cc/PV9M-CKCH] ("While deepfakes have received enormous attention for their potential political dangers, the vast majority of them are used to target women. Sensity AI, a research company that has tracked online deepfake videos since December of 2018, has consistently found that between 90% and 95% of them are nonconsensual porn.").

[11] Kristen Dold, *Face-Swapping Porn: How a Creepy Internet Trend Could Threaten Democracy*, ROLLING STONE (Apr. 17, 2018, 8:47 PM), https://www.rollingstone.com/culture/culture-features/face-swapping-porn-how-a-creepy-internet-trend-could-threaten-democracy-629275/ [https://perma.cc/AS7F-TQDB]; Damon Beres & Marcus Gilmer, *A Guide to 'Deepfakes,' the Internet's Latest Moral Crisis*, MASHABLE (Feb. 2, 2018), https://mashable.com/2018/02/02/what-are-deepfakes/#FPVRcf.91qqM [https://perma.cc/KE23-4RNE]; *Yes, Positive Deepfake Examples Exist*, THINK AUTOMATION, https://www.thinkautomation.com/bots-and-ai/yes-positive-deepfake-examples-exist/ [https://perma.cc/VR2Y-HT6P].

[12] Beres & Gilmer, *supra* note 11.

[13] Meredith Somers, *Deepfakes, Explained,* MIT MGMT. SLOAN SCH. (July 21, 2020), https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained [https://perma.cc/G2EP-DVXJ].

[14] Raina Davis, Chris Wiggins & Joan Donovan, *Deepfakes*, HARV. KENNEDY SCH. BELFER CTR. FOR SCI. AND INT'L AFF. 3, https://www.belfercenter.org/sites/default/files/files/publication/Deepfakes_2.pdf [https://perma.cc/2AY8-KM47].

[15] *See id.*

The first method utilizes generative adversarial networks ("GANs").[16] A pair of GANs are used in which one GAN network "inputs a latent sample and generates an image."[17] Then, in simplified terms, the output created from the first network is fed into a second network that identifies the image as genuine or fake.[18] The network continually provides feedback to itself, rating the image on a probability scale weighing the authenticity of the image while constantly adjusting the image until the computer itself can no longer determine the difference between the original input image and the new deepfake that has emerged.[19]

The alternative method of creating a deepfake involves utilizing deep learning computers called variational auto-encoders ("VAEs").[20] VAEs include two networks working together that are trained and programmed to encode an image into a low-dimensional representation and then subsequently decode the representation back into an image.[21] The decoder houses hundreds of images of a desired celebrity or public figure that are continuously configured until the input and output images match.[22] The decoder then acts analogous to an artist, and it can add additional features like hats, sunglasses, or accessories.[23] In other words, if the objective was to create a realistic looking video of the late former President George Washington riding a Segway, one encoder would be trained and programmed to identify images of George Washington's face while the other would be trained on a wide variety of alternative faces.[24] The images of the faces identified by the encoder can be curated to create different frames, poses, and lighting.[25] Once the encoder network training is complete, the output from both encoders are combined, resulting in George Washington's face on someone else's body; that body just happens to be riding a Segway or partaking in another humorous, sensational, or malevolent act.[26]

Obviously, since George Washington has passed, it would be quite

---

[16] *Id.*

[17] *Id.*

[18] Naoki Shibuya, *Understanding Generative Adversarial Networks,* MEDIUM (Nov. 2, 2017), https://medium.com/activating-robotic-minds/understanding-generative-adversarial-networks-4dafc963f2ef [https://perma.cc/MF2L-3XBX].

[19] *Id.*

[20] *Artificial Intelligence: GANs and Autoencoders Applied to CyberSecurity,* ELEVEN PATHS 10 (May 30, 2019), https://pro-cdo-web-resources.s3.eu-west-1.amazonaws.com/elevenpaths/uploads/2020/6/elevenpaths-whitepaper-artificial-intelligence-gans-and-autoencoders-applied-to-cybersecurity.pdf [https://perma.cc/EHA5-FYR5].

[21] Davis, et al., *supra* note 14.

[22] *Id.*

[23] *Id.*

[24] Somers, *supra* note 13.

[25] *Id.*

[26] *Id.*

easy to identify the video as fake.[27] However, in general, deepfakes can be quite difficult to identify as inauthentic because they generally use genuine footage, crisp audio, and are popularly shared to hundreds of thousands, if not millions of people via social networks such as Twitter, Facebook, Instagram, and the newly prevalent TikTok.[28] Additionally, with the mainstream prevalence of "fake news,"[29] deepfakes will most likely only fuel the fake news fire by allowing users to provide misleading information in a variety of convincing ways.[30]

Deepfakes have the potential to affect unlimited facets of our daily lives including personal relationships, education and professional opportunities, the political realm, and the legal field. In some aspect, their use may create positive applications for an efficient and edified life, which will be discussed later.[31] However, their mere existence could lead to a bleak outcome if the population is ill-informed as to their existence. The legal field in particular must understand deepfakes' potential harm and unethical behavior that stems from them.

III. THE SOCIAL WORLD AND DEEPFAKES

Deepfakes, whether you realize it or not, are already a significant component of our social lives because their use increases daily in social media, art, politics, and more.[32] Aside from the troubling sexualization aspects of deepfake technology, Hollywood, social media, and the political realm alike have employed deepfakes for entertainment, education, and experimentation.[33] When used improperly, deepfakes can have a damaging

---

[27] Blanton, Wyndham, *The Death of George Washington*, GEORGE WASHINGTON'S MOUNT VERNON, https://www.mountvernon.org/library/digitalhistory/digital-encyclopedia/article/the-death-of-george-washington/ [https://perma.cc/V2AA-7WTW].

[28] Mika Westerlund, *The Emergence of Deepfake Technology: A Review*, 9 TECH. INNOVATION MGMT. REV. 11, 40 (Nov. 2019), https://timreview.ca/article/1282 [https://perma.cc/Q4B6-H27C].

[29] *What is Fake News?*, GOODWILL CMTY. FOUND., https://edu.gcfglobal.org/en/thenow/what-is-fake-news/1/ [https://perma.cc/2LEW-U8EN].

[30] Oscar Schwartz, *You Thought Fake News Was Bad? Deep Fakes Are Where Truth Goes to Die*, THE GUARDIAN (Nov. 12, 2018), https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth [https://perma.cc/T6UY-9YEV].

[31] Yisroel Mirsky & Wenke Lee, *The Creation and Detection of Deepfakes: A Survey*, 1 ACM COMPUT. SURV. 1 (Jan. 2020), https://arxiv.org/pdf/2004.11138.pdf [https://perma.cc/7MGQ-46ZH].

[32] *See* John Brandon, *There Are Now 15,000 Deepfake Videos on Social Media. Yes, You Should Worry.*, FORBES (Oct. 8, 2019), https://www.forbes.com/sites/johnbbrandon/2019/10/08/there-are-now-15000-deepfake-videos-on-social-media-yes-you-should-worry/?sh=1691fdc23750 [https://perma.cc/5PDQ-KL6E].

[33] Jeremy Kahn, *Forget Disinformation. It's Hollywood and Madison Avenue Where Deepfakes Are About to Wreak Havoc*, FORTUNE (June 22, 2021),

effect.[34] However, examples of positive uses of deepfakes and their technology have already emerged.[35]

Deepfake technology has wide potential for developments in education and social change. In 2018, the Illinois Holocaust Museum and Education Center launched an exhibit in which museum patrons could interact with hologrammatic interviews of Holocaust survivors by asking them questions and hearing their stories.[36] Similarly, a tech company resurrected the late John F. Kennedy's voice and used artificial intelligence ("AI") deepfake technology to allow his "voice" to deliver the speech he would have delivered, if not for his assassination.[37] Finally, deepfake technology can assist in conquering language barriers.[38] In a public service announcement advocating for awareness and finding a cure for malaria, deepfake AI technology was used to make David Beckham speak multiple languages in order to spread the advantageous message to various countries in their native languages.[39]

In a less enriching example, the late Carrie Fisher was brought back to life using deepfake technology to reprise her role as Princess Leia in the latest Star Wars film.[40] Most notably in late 2021, multiple deepfakes of famous action star Tom Cruise have been making rounds on social media sites like TikTok and Facebook.[41] Staying true to a hallmark of deepfakes in that the subject is often partaking in uncharacteristic actions, the AI Cruise is seen performing a magic trick, eating a lollipop, and singing a song.[42] Referring to the Tom Cruise deepfakes, Hany Farid, a professor at the University of California, Berkeley, told National Public Radio that "this

---

https://fortune.com/2021/06/22/deepfakes-tom-cruise-chris-ume-metaphysic-hollywood-madison-avenue-eye-on-ai/ [https://perma.cc/QRY3-RPPN]; Brandon, *supra* note 32; Leonie Bos, *Deepfakes Are Jumping from Porn to Politics. It's Time to Fight Back*, WIRED (Dec. 20, 2020), https://www.wired.co.uk/article/deepfakes-porn-politics [https://perma.cc/MW5M-PRLK].

[34] Scott, *supra* note 6 ("I had spent my entire adult life watching helplessly as my image was used against me by men that I had never given permission to of any kind.").

[35] Ashish Jaiman, *Positive Use Cases of Deepfakes*, TOWARDS DATA SCI. (Aug. 14, 2020), https://towardsdatascience.com/positive-use-cases-of-deepfakes-49f510056387 [https://perma.cc/D45V-EGMM].

[36] THINK AUTOMATION, *supra* note 11.

[37] *Id.*

[38] *Id.*

[39] *Id.*

[40] Joseph Foley, *14 Deepfake Examples That Terrified and Amused the Internet*, CREATIVE BLOQ (June 1, 2021), https://www.creativebloq.com/features/deepfake-examples [https://perma.cc/WH3E-RM9P].

[41] Rachel Metz, *How a Deepfake Tom Cruise on TikTok Turned into a Very Real AI Company,* CNN BUS. (Aug. 6, 2021), https://www.cnn.com/2021/08/06/tech/tom-cruise-deepfake-tiktok-company/index.html [https://perma.cc/5GLT-FJGL].

[42] *Id.*

is clearly a new category of deepfake that we have not seen before."[43] Farid, who researches digital forensics and misinformation, pronounced that the Cruise deepfakes demonstrate a "step up" in the technology's developing advancement.[44]

Although there are numerous positive aspects to the rapidly evolving technology of deepfakes, their presence unearths a growing concern of their uses, including fostering an environment that lacks trust along with the potential for hackers and scammers to exploit vulnerable and naïve citizens. In response to the Tom Cruise deepfakes, former Central Intelligence Agency officer and disinformation specialist Matt Ferraro indicated that the national security and intelligence community in Washington D.C. now believes we are one step closer to a feared dystopian future.[45] When referring to the security community, Ferraro prophesied the source of fear in the community by saying, "It's because they realize how dangerous [deepfakes] are. It does seem like it's really going to be a fundamental challenge to the information environment."[46]

Moreover, the nightly broadcast of a popular news program may be shrouded in doubt as deepfake videos of politicians like former Presidents Barack Obama and Donald Trump are shared, only later found to be created with deepfake technology.[47] Also, audio deepfakes have already been used to clone voices and convince people they are speaking to trusted companions, convincing them to partake in activity they normally would not.[48] Overall, deepfakes have the potential to erode and undermine our trust in everyday life. Although they create new opportunities for learning and education,[49] things will not always be as they seem with the advent and

---

[43] Emma Bowman, *Slick Tom Cruise Deepfakes Signal That Near Flawless Forgeries May Be Here*, NPR (Mar. 11, 2021, 5:47 PM), https://www.npr.org/2021/03/11/975849508/slick-tom-cruise-deepfakes-signal-that-near-flawless-forgeries-may-be-here [https://perma.cc/H9KR-EERZ].

[44] *Id.*

[45] Mark Corcoran & Matt Henry, *The Tom Cruise Deepfake that Set Off 'Terror' in the Heart of Washington DC*, ABC NEWS AUSTL. (June 27, 2021, 7:16 PM), https://www.abc.net.au/news/2021-06-24/tom-cruise-deepfake-chris-ume-security-washington-dc/100234772 [https://perma.cc/XLM9-5JV3].

[46] *Id.*

[47] Kaylee Fagan, *A Viral Video that Appeared to Show Obama Calling Trump a 'Dips---' Shows a Disturbing New Trend Called 'Deepfakes,'* INSIDER (Apr. 17, 2018, 3:48 PM), https://www.businessinsider.com/obama-deepfake-video-insulting-trump-2018-4 [https://perma.cc/GG3A-8E5K].

[48] Jesse Damiani, *A Voice Deepfake Was Used to Scam a CEO Out of $243,000*, FORBES (Sept. 3, 2019), https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=62d555282241 [https://perma.cc/S5XF-A3V5].

[49] Matthew Griffin, *Edtech Company Udacity Uses Deepfake Tech to Create Educational Videos Automatically*, FANATICAL FUTURIST (Aug. 2, 2019), https://www.fanaticalfuturist.com/2019/08/edtech-company-udacity-uses-deepfake-tech-to-create-educational-videos-automatically/ [https://perma.cc/AD3V-Z5BC] ("Producing

ease of using deepfake technology.

IV. THE EVOLUTION OF EVIDENCE

Indubitably, the courtroom and legal system will be meaningfully affected by deepfake technology. The most glaring application of deepfake technology to the legal system is its entry in generating unrest and inconsistency in the accusation of crime, dispute settlements, and the courtroom itself, including *Daubert* hearings on evidence, which will be discussed further on.[50] Prior to the prevalence of text messages, videos, and emails, eyewitness accounts were the pinnacle of evidentiary support, often decimating the high and mighty jury from its pre-conceived notions of the case at hand, forcing it to consider alternatives to its gut theory.[51]

In the early eighteenth century, the preferred method of evidence proffered to the court was written.[52] Written proof was favored over oral testimony because at this time, oral or unwritten evidence that was submitted became entangled with numerous criteria that are not present today, including religious disabilities, lack of discernment, incompetency, conflicts of interest, or blatant inability to testify for a variety of causes.[53] However, in 1755, written testimony was often shrouded in confusion and conflict, while oral testimony appeared to follow with less controversy.[54] This may be due to the fact that hearsay restrictions, when present, were almost never regulated while expert testimony parameters were extremely loose in comparison to the rigid standards of today.[55]

In the late nineteenth century and early into the twentieth century, the reign of oral evidence continued.[56] Documents were considered unreliable and confusing to the jury.[57] In an early treatise on evidence, Thomas Starkie, who is considered a founder of the modern evidentiary argument, stated:

Oral testimony, it is to be remarked, in natural order precedes

---

content for Massive Open Online Course (MOOC) platforms like Coursera and EdX might be academically rewarding, and potentially lucrative, but it's also hugely time consuming – particularly where videos are involved, so Udacity, in an ode to Soul Machines who recently created 'Will' the world's first avatar teacher who's already taught over 250,000 children about energy, have been looking into ways to get Artificial Intelligence (AI) to produce the videos automatically for them – something that would be a game changer in the academic world.").

[50] *See infra* Part VIII.

[51] *See* John H. Langbein, *Historical Foundations of the Law of Evidence: A View from the Ryder Sources*, 96 COLUM. L. REV. 1168 (1996) (explaining how dependence on witness testimony shaped evidentiary law).

[52] *Id.* at 1183.

[53] *Id.* at 1194–95.

[54] Thomas P. Gallanis, *The Rise of Modern Evidence Law*, 84 IOWA L. REV. 499, 530 (1999).

[55] *Id.* at 513.

[56] *See id.* at 529.

[57] *See id.* (noting oral evidence was more common than written evidence in criminal trials during the early 1820s).

written evidence. It is in general more proximate to the fact than written evidence, being a direct communication by one who possesses actual knowledge of the fact by his senses; whilst written evidence in itself requires proof, and must ultimately be derived from the same source with oral evidence, that is, from those who possessed actual knowledge of the facts.[58]

However, as the realm of evidence continued to evolve, empirical research about the topic of oral testimony and in particular, memory, commenced.[59] With that, reservations about the reliability of oral testimony started breaching the evidentiary surface.[60]

As an example, in 1901 in Berlin, a criminal law professor was unremarkably lecturing to his law students when abruptly, a student shouted an objection to his line of argument.[61] In response, an angered student hurled corrosive insults, and a heated altercation began between the students.[62] Finally, the valiant professor intervened and calmed the situation down.[63] Little did the audience of scared law students know, the entire scene was a staged exercise procured to test the strength and vigor of their recollection of the events.[64] Not surprisingly, the results were abysmal with the most student witnesses recalling twenty-five to eighty percent of the significant details wrong.[65] Pivotal moments in the scene were completely wiped away from the students' memories while others described events that only took place in their lucid imaginations.[66]

In a subsequent 1932 study, Sir Frederic Charles Bartlett conducted an experiment now known as "The War of the Ghosts."[67] This study demonstrated the fragile nature of memory by assigning English participants to read and remember a Canadian Indian Folklore story.[68] When instructed to recount portions of the folklore, the English participants' recollection was skewed; readers often replaced unfamiliar or culturally different

---

[58] THOMAS STARKIE, GEORGE MORLEY DOWDESWELL & JOHN GEORGE MALCOLM, PRACTICAL TREATISE OF THE LAW OF EVIDENCE 103 (8th American ed., Philadelphia, T. & J.W. Johnson 1860).

[59] *See generally* Alice M. Hoffman & Howard S. Hoffman, *Reliability and Validity in Oral History: The Case for Memory*, *in* MEMORY AND HISTORY: ESSAYS ON RECALLING AND INTERPRETING EXPERIENCE 107, 108 (Glenace E. Edwall & Jaclyn Jeffrey eds., 1994).

[60] *See generally id.*

[61] Atul Gawande, *Investigations Under Suspicion: The Fugitive Science of Criminal Justice*, NEW YORKER, Jan. 8, 2011, at 50.

[62] *Id.*

[63] *Id.*

[64] *Id.*

[65] *Id.*

[66] *Id.*

[67] Harald Sack, *Frederic Bartlett and Experimental Psychology*, SCIHI BLOG (Oct. 20, 2016), http://scihi.org/frederic-bartlett-and-experimental-psychology/ [https://perma.cc/45SW-FUSW].

[68] *Id.*

information with something more familiar.[69] For instance, participants changed a mentioned activity of seal hunting to fishing, and the use of canoes to regular boats.[70] The readers also postulated multiple inferences pertaining to the story that went beyond the scope of the information originally provided in the tale.[71] The ever-evolving distrust in oral witness testimony paves a clear and relatively unobstructed path for reliance and confidence in video forms of evidence, particularly in the courtroom.

V. DEEPFAKES AND THE COURT SYSTEM

Shifting abruptly to our current era, the technology landscape we live in is rapidly expanding due to the availability and ease of using portable and potent video and audio capturing technology, like police body cameras, smart phones, and dashboard recording devices. Even prior to the new age in technology, legal commentators have noted that it is "both reasonable and necessary for the legal profession to improve its procedures by taking advantage of important new technologies such as videotape."[72] The court system has widely adopted the use of technological evidence, and the vast increase in the quality and quantity of digital evidence has been a challenge for the court system in terms of receiving, authenticating, and presenting video evidence to the court.[73] Now, with the mass influx of new technology and the rise of deepfakes, the court system is presented with new challenges that may once have never been pondered.[74] Obviously, the main remonstrance today focuses on the identification of altered and fraudulent deepfake videos as well as stopping them in their tracks before the use of them in a legal proceeding radically alters the outcome.[75]

"To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a

---

[69] William F. Brewer, *Learning Theory Schema Theory*, EDUC. ENCYCLOPEDIA, https://education.stateuniversity.com/pages/2175/Learning-Theory-SCHEMA-THEORY.html [https://perma.cc/9ULQ-CB6A].

[70] *Bartlett's 'War of the Ghosts'*, PSYCHOLOGY GCSE, https://sites.google.com/view/gcsepsychology/cognition-and-behaviour/bartletts-war-of-the-ghosts [https://perma.cc/Y63H-EV9T].

[71] *Id.*

[72] 44 AM. JURIS. TRIALS 171 (originally published in 1992).

[73] *See* JOINT TECH. COMM., NAT'L CTR. FOR STATE CTS., JTC RESOURCE BULLETIN: MANAGING DIGITAL EVIDENCE IN COURTS, (2016), https://www.ncsc.org/__data/assets/pdf_file/0017/18521/digital-evidence-3-14-2016-final.pdf [https://perma.cc/44SY-DT8W].

[74] Matt Reynolds, *Courts and Lawyers Struggle with Growing Prevalence of Deepfakes,* A.B.A. J. (June 9, 2020), https://www.abajournal.com/web/article/courts-and-lawyers-struggle-with-growing-prevalence-of-deepfakes [https://perma.cc/TW29-T33A].

[75] *See* Esquire Deposition Solutions, LLC, *Protecting Legal Proceedings from Deepfake Technology,* JD SUPRA (Feb. 18, 2020), https://www.jdsupra.com/legalnews/protecting-legal-proceedings-from-90642/ [https://perma.cc/FKF7-Q92J].

finding that the item is what the proponent claims it is."[76] Specifically, for photographs and videos, either party can authenticate the visual depiction as a "fair and accurate representation of the underlying scene at the relevant time."[77] However, courts have inferred that the bar for authentication of evidence is "not particularly high," and the Federal Rule of Evidence for authentication of video only requires that a reasonable jury could find in favor of authenticating the visual.[78] In other words, the jury will ultimately decide if the visual is what it purports to be, but the initial threshold of admitting it is not a difficult hurdle to overcome.[79]

Federal Rule of Evidence 902(13) calls for authentication of records "generated by an electronic process or system that produces an accurate result" if "shown by a certification of a qualified person" in a particular way.[80] However, due to the sheer, real quality of the videos, witnesses or proponents of the video could have been psychologically deceived or fooled into believing the reality of the video.[81] In more pessimistic terms, the so-called qualified person testifying that the video is what it says to be could simply be lying.[82] It is important to note that based on the Federal Rules of Evidence, just because a video or visual is admissible under authentication scrutiny does not mean it is automatically admissible.[83] The purported evidence may still be deemed inadmissible under any of the other applicable Federal Rules of Evidence if it does not meet the additional requirements of that specific rule.[84]

VI. THE COURTROOM AND EVIDENCE AUTHENTICATION

Throughout the years, various court holdings have demonstrated that in terms of evidence authentication and admissibility, there is a relatively

---

[76] FED. R. EVID. 901.

[77] DEBORAH JONES MERRITT & RIC SIMMONS, LEARNING EVIDENCE: FROM THE FEDERAL RULES OF THE COURTROOM, 919–24 (4th ed. 2018).

[78] FED. R. EVID. 901(a); *see* United States v. Vayner, 769 F.3d 125, 130 (2d Cir. 2014) ("The ultimate determination as to whether the evidence is, in fact, what its proponent claims is thereafter a matter for the jury.").

[79] United States v. Gammal, No. 19-468, 2020 U.S. App. LEXIS 33913, at *2–3 (2d Cir. Oct. 26, 2020).

[80] FED. R. EVID. 902(13); *see* Theodore F. Claypoole, *AI and Evidence: Let's Start to Worry,* NAT. L. REV. (Nov. 14, 2019), https://www.natlawreview.com/article/ai-and-evidence-let-s-start-to-worry [https://perma.cc/W93F-BFDY].

[81] *See* Claypoole, *supra* note 80.

[82] *Id.*

[83] *See* Richards v. McClure, 858 N.W.2d 841, 849 (Neb. 2015) (citing 29A AM. JUR. 2D *Evidence* § 1048 at 389 (2008)) (explaining that even if a document is authenticated as required for admissibility, that authentication does not automatically guarantee admissibility).

[84] *See* MERRITT & SIMMONS, *supra* note 77, at 919–24.

low threshold.[85] Specifically, this perception rings true when a witness or party can purportedly identify and confirm that an evidentiary image or video is accurate.[86] Back in 1985, in a case when no well-founded accusation of inaccuracy was made, the simple testimony of an investigator that a videotape showed accurate depictions of an event tantamount to the trial was sufficient for video authentication purposes.[87]

In a 2010 California appellate court decision, although no expert testified that a photograph was "composite" or "faked," the court ruled the prosecutor should not have admitted a Myspace photo claiming to show the girlfriend of a defendant flashing a gang sign because the photo had not been authenticated.[88] Although the court ruled the error did not cause harm to the defendants, the court mentioned that with the "advent of computer software programs such as Adobe Photoshop, 'it does not always take skill, experience, or even cognizance to alter a digital photo.'"[89] In 2016, DVD recordings containing surveillance footage were properly authenticated by a witness who had knowledge of the company and understood how the system worked, including the display of the date and time on the footage.[90]

More recently in 2021, an appellate court considered the challenges of evidence brought about by new technology by stating, "[W]e are mindful that in the age of fake social-media accounts, hacked accounts, and so-called deepfakes, a trial court faced with the question whether a social-media account is authentic must itself be mindful of these concerns."[91] In this case, the court's decision to admit evidence was questioned on appeal regarding the authentication of Facebook posts.[92] An intended person had made ill and malicious remarks during a Facebook Live video, so as revenge, the defendant shot a person he believed to be the curator of the Facebook Live video.[93] In its reasoning, the court noted concerns about fake social media accounts but conditionally authenticated proffered social media evidence with the option for objection by an opponent if further evidence of

---

[85] *See infra* notes 86–90.

[86] George L. Blum, Annotation, *Authentication of Social Media Records and Communications*, 40 A.L.R. 7th Art. 1 (2019) (citing State v. Patterson, No. C-170329, 2018 WL 4026476, at *3 (Ohio Ct. App. Aug. 22, 2018)) (testimony that victim saw defendant's photograph on Facebook was sufficient to meet the low threshold required to authenticate the photograph).

[87] Louis Vuitton S.A. v. Spencer Handbags Corp., 765 F.2d 966, 974 (2d Cir. 1985).

[88] People v. Beckley, 185 Cal. App. 4th 509, 515 (2010).

[89] *Id.* (quoting 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, 183 (2009)).

[90] United States v. Kessinger, No. 15-5364, 2016 U.S. App. LEXIS 2408, at *11 (6th Cir. Feb. 9, 2016).

[91] People v. Smith, No. 346044, 2021 Mich. App. LEXIS 1144, at *34 (Ct. App. Feb. 18, 2021).

[92] *Id.* at *2.

[93] *Id.*

deception arose.[94] However, the court opined that the decision in this case "does not discount the possibility that evidence from social media, might in fact, be inaccurate, hacked, or faked. As technology advances, trial courts and lawyers will need to be vigilant when considering questions of authenticity, at both the first and second stages."[95]

Alternatively, in 2015, a separate California jurisdiction disagreed with *Beckley* in which the court articulated that the foundation for a photograph or video "'may be supplied by other witness testimony, circumstantial evidence, content and location' and 'may also be established "by any other means provided by law" including a statutory presumption.'"[96] The court stated that "reading *Beckley* as equating authentication with proving genuineness would ignore a fundamental principle underlying authentication."[97] This principle is that the ultimate determination of authenticity of evidence is for the trier of fact.[98] The trier of fact is in the position to consider the evidence presented to them and weigh it against any inconsistencies to eventually arrive at a final determination on authenticity.[99] Likewise, when a plaintiff could not detail how a video was made or if it had been altered, the Seventh Circuit Court of Appeals did not allow the evidence to be authenticated.[100] In doing so, the court properly mirrored the California Court of Appeals in identifying fraudulent material.[101]

An appellate court in 2019 held that the trial court did not abuse its discretion in admitting a voicemail professedly left by defendant for the murder victim.[102] In *Gonzales*, the court looked to the rulings of many jurisdictions and adopted a flexible approach to authenticating video and photo recordings.[103] However, the court noted that computer technology and software are in general public use, and most owners of a computer have the requisite knowledge and programs to potentially tamper with recordings.[104] Despite the court's awareness of the potential for deception, it reasoned that the possibility alone is not enough to postulate narrow and restrictive rules for authentication that must be applied in every case, specifically if there is no corroborative reason to do so.[105]

The Colorado Court of Appeals in *Gonzales*, like other jurisdictions,

---

[94] *Id.* at *35.
[95] *Id.* at *38.
[96] *In re* K.B., 238 Cal. App. 4th 989, 995 (2015) (quoting People v. Goldsmith, 59 Cal. 4th 258, 268 (2014)) (citations omitted).
[97] *Id.* at 997.
[98] *Id.*
[99] *Id.*
[100] Griffin v. Bell, 694 F.3d 817, 826 (7th Cir. 2012).
[101] *Bell*, 694 F.3d at 826; *In re* K.B., 238 Cal. App. 4th at 998.
[102] People v. Gonzales, 474 P.3d 124, 130–31 (Colo. App. 2019).
[103] *Id.* at 130.
[104] *Id.*
[105] *Id.*

looked to *People v. Sangster* for guidance.[106] In *Sangster,* the state sought to introduce evidence of a telephone call apparently made by the defendant from jail to an unnamed female.[107] The defendant in this case was allegedly using the jail telephone system to contact potential witnesses in his upcoming case with the intent to coerce or threaten them into absenteeism.[108] When inmates entered the population of this jail, they were registered into the jail's telephone system.[109] The system consisted of a voice recognition component coupled with the use of a personal identification number.[110] When an inmate attempted to make a call, they were required to provide their personal identification number and audibly state their name.[111] If the inmate's voice corresponded to their previously recorded voice, the call was dispatched.[112] However, if the voice associated with the specific personal identification number did not match, the call would disconnect.[113]

The defendant's claim that the phone call evidence was improperly admitted fell short.[114] In this First District Court of Appeals case in Illinois, the court posited that when a defendant does "not present any actual evidence of tampering, substitution, or contamination, the State need only establish a probability that those things did not occur."[115] As many skilled trial lawyers often opine, "any deficiencies go to the weight, rather than the admissibility, of the evidence."[116]

Similarly, again citing *Sangster,* earlier on in 2020, a defendant's argument failed when the defendant argued that "improperly-authenticated recordings are inherently suspect in this age of deep-fake videos and easily-manipulated audio records."[117] In this case, the court paralleled the *Sangster* reasoning and postulated that unless there is apparent evidence of tampering or manipulation, there may not be foundational issues with the evidence presented.[118]

Clearly infiltrating the United States' court system, deepfakes also have a global reach in worldwide courts. In an ongoing United Kingdom child custody battle, a wife submitted into evidence a heavily doctored audio recording of the children's father in which the father was making violent and

---

[106] *Id.*
[107] People v. Sangster, 8 N.E.3d 1116, 1124 (Ill. App. Ct. 2014).
[108] *Id.*
[109] *Id.*
[110] *Id.*
[111] *Id.*
[112] *Id.*
[113] *Id.*
[114] *Id.* at 1128.
[115] *Id.* at 1127–28.
[116] *Id.* at 1128.
[117] People v. Foreman, No. 2-18-0178, 2020 Ill. App. LEXIS 1240, at *54 (Ill. App. Ct. July 17, 2020), *appeal denied*, 159 N.E.3d 936 (Ill. 2020).
[118] *Id.* at *55.

malicious threats towards his wife and children.[119] Luckily, experts were able to identify the manipulated recording.[120] However, the doctored clip in this case was a less sophisticated version of deepfake AI videos and is suitably called "cheapfake" technology.[121] Mr. James, a partner in the London-based law firm Expatriate Law, suggested that "it would never occur to most judges that deepfake material could be submitted as evidence."[122]

These global cases appear to show that some courts are in fact aware of the existence and potential issues caused by deepfake technology. However, as demonstrated in the various approaches to authenticating evidence throughout United States' jurisdictions, there is unconcealed leniency in authenticating evidence unless it can be proved that there is reason to believe of foul play in modifying or editing an auditory or visual piece of evidence.[123] The prevalence of deepfakes results in the need for a call to action for all participants in the legal community to acknowledge the threat, educate themselves on methods of detection, and advocate for others to become aware of the impending technological threat. Deepfakes have the potential to cause an extreme erosion of trust in the public perception of the courtroom.[124]

VII. COMBATTING DEEPFAKES

In order to deter the use of deepfake technology, proponents of all facets in the legal community need to be educated and aware of this issue and recognize that additional resources and investments may be needed to properly battle their infiltration into the legal system.[125] For instance, if a client is pushing evidence upon a lawyer with suspect motives, it may be a red flag that the evidence is less than authentic. Additionally, training in spotting outward signs of altered deepfake technology can be a basic, but needed, starting point in weeding out the malignant deepfake tumors in the

---

[119] *See* Reynolds, *supra* note 74.

[120] *Id.*

[121] *Id.*

[122] Patrick Ryan, *'Deepfake' Audio Evidence Used in UK Court to Discredit Dubai Dad*, NAT'L NEWS: UAE (Feb. 8, 2020), https://www.thenationalnews.com/uae/courts/deepfake-audio-evidence-used-in-uk-court-to-discredit-dubai-dad-1.975764 [https://perma.cc/Y9EU-HAG4].

[123] *See generally supra* Part VI.

[124] *Cf.* Kelsey Farish, *The Legal Implications and Challenges of Deepfakes*, DAC BEACHCROFT (Sept. 4, 2020), https://www.dacbeachcroft.com/en/gb/articles/2020/september/the-legal-implications-and-challenges-of-deepfakes/ [https://perma.cc/QD8Z-M8JX] (briefly observing deepfakes can be used in creating fraudulent submissions of documentary evidence while providing examples in which deepfakes present risks to various aspects of our socio-political system).

[125] *See* People v. Smith, No. 346044, 2021 Mich. App. LEXIS 1144, at *38 (Mich. Ct. App. Feb. 18, 2021), *appeal denied*, 962 N.W.2d 277 (Mich. 2021).

court system.[126] Overall, countermeasures throughout the legal system need to be enacted to detect and prevent the use of mischievous deepfake technology.[127]

In some jurisdictions, courts continue to rely on less sophisticated methods of identification as a first and minor step in identifying these manipulated videos when technological methods may not be easily accessible.[128] This includes using a totality of the circumstances approach in examining if the actions, words, and mannerisms of the supposed individual on video is acting in accordance with their known personality.[129] Additionally, less complex identification methods include gauging promises or comments made by a deepfake video and simply seeing if those promises, threats, or concerns come to fruition.[130] For example, an executive in the United Kingdom was shown a video of his chief executive officer requesting a wire transfer to an unknown account in a large sum of money.[131] The executive only learned of the deceit because the reimbursement of the transferred funds promised by the CEO never came.[132] Although utilizing the notions of common sense and caution are useful tactics, as deepfakes become more prevalent in society and the courtroom, technology involving their detection will become indispensable in a courtroom scenario.[133]

The easiest, albeit not the most advanced, technological solution to detecting deepfakes is to conduct a reverse-image search.[134] "Reverse image search is a search engine technology that takes an image file as input query and returns results related to the image."[135] Search engines like Google allow a user to navigate to the search image page and upload an image from the

---

[126] *Cf. id.* at 34–39 (advocating for deepfake detection education and training as a starting point for alleviating potential courtroom-based manipulation based on the suggestion that trial courts need to be mindful of concerns surrounding fake social media and deepfakes).

[127] Mirsky & Lee, *supra* note 31, at 1:26–28.

[128] Christy Foster, *Deepfake Evidence – What It Is and How to Spot It,* LEXOLOGY (Feb. 13, 2020), https://www.lexology.com/library/detail.aspx?g=be75a3a5-595b-4dc8-ac4e-7e6f0523257f [https://perma.cc/H5NC-KM7S] (noting that although rare, members of the legal community need to be apprised of deepfakes in the legal field as technology continues to improve).

[129] *See id.*

[130] *Id.*

[131] *Id.*

[132] *Id.*

[133] *Cf. id.* (describing the idea that there currently is not a "one size fits all" detection method for deepfakes and as technology continues to improve, legal advisors and litigants alike should consider being aware of their presence).

[134] Nicholas Schmidt, *Privacy Law and Resolving 'Deepfakes' Online*, IAPP (Jan. 30, 2019), https://iapp.org/news/a/privacy-law-and-resolving-deepfakes-online/ [https://perma.cc/7JKY-RVG4].

[135] Ivy Wigmore, *Reverse Image Search*, TECH TARGET (Apr. 2013), https://whatis.techtarget.com/definition/reverse-image-search [https://perma.cc/N2LU-QX4L].

user's computer or a specific URL link.[136] From there, the search engine will scour the internet to assist in locating the source for an image, searching for duplicated content, debunking faked images, and ensuring compliance with copyright regulations.[137] Using this technique, if a deepfake was created by another image or video already located somewhere on the internet, the original version should appear in the reverse image search.[138]

In a more abstract detection method, deepfakes are potentially spotted by a scrupulous study of the image employing the "uncanny valley" method.[139] First coined in the 1970s when a Japanese roboticist Masahiro Mori conducted an experiment using a lifelike robot, the "uncanny valley" hypothesis asserts that when humans are faced with images, videos, or even the physical presence of an artificial figure, the human appearance and behavior of the entity can make a spectator believe the entity is human.[140] However, "the sense of viewer familiarity drops into an 'uncanny valley' once the artificial figure tries but fails to mimic a realistic human."[141]

Since empirical data is inconsistent on the issue, the "uncanny valley" hypothesis is currently still open to debate, but it appears to gather support from evolutionary, social, cognitive, and psychodynamic viewpoints.[142] Because "[d]eepfakes are often assembled by an algorithm from still photos," natural human facial expressions and common but unconscious bodily ticks are often not apparent in the final product, and if they are, they will not fully translate, seeming odd or strange in comparison to natural human movement and activity.[143] In short, if following the "uncanny valley" hypothesis, deepfake videos should evoke a "subconscious feeling of unease" and deceit, often triggering the viewer to question the validity of the content.[144]

Turning to a drastically more complex detection method for deepfakes, organizations like the Massachusetts Institute of Technology and the U.S. Department of Defense are using biometric data to identify inconsistencies in video data.[145] Biometrics are the unique physical

---

[136] *Id.*

[137] *Id.*

[138] Schmidt, *supra* note 134.

[139] *Id.*

[140] Jeremy Hsu, *Why "Uncanny Valley" Human Look-Alikes Put Us on Edge*, SCI. AM.: INNOVATIONNEWSDAILY (Apr. 3, 2012), https://www.scientificamerican.com/article/why-uncanny-valley-human-look-alikes-put-us-on-edge/ [https://perma.cc/4FGQ-KEHP]; Masahiro Mori, *The Uncanny Valley*, IEEE SPECTRUM (June 12, 2012), https://spectrum.ieee.org/the-uncanny-valley [https://perma.cc/4U8Q-B2T5].

[141] *Id.*

[142] *See* Shensheng Wang, Scott O. Lilienfeld & Philippe Rochat, *The Uncanny Valley: Existence and Explanations*, 19 REV. OF GEN. PSYCH. 393, 395–98 (2015).

[143] Schmidt, *supra* note 134.

[144] *See id.*

[145] *See* Satya Venneti, *Real-Time Extraction of Biometric Data from Video*, CARNEGIE

characteristics of a person, like their fingerprints, that can be measured and used for automated recognition.[146] One method of employing biometric data to defeat deepfakes involves iris scanning.[147] An iris, which is made of connective tissue and smooth muscle fibers within the eye, is responsible for working in conjunction with the pupil of an eye to regulate light.[148] Thus, to identify a potential deepfake, a scan of a subject's iris in a video can reveal pupil size and dilation.[149] Any deviations from the norm could be cause to investigate further, potentially exposing a deepfake.[150] In fact, the now famous Tom Cruise deepfakes were exposed as fraudulent partially because experts noticed Cruise's pupils were distorted.[151]

Likewise, biometric data can be used to identify deepfake videos when the blood flow of a subject in a video is analyzed.[152] In reality, actual humans exhibit similar blood flow and pulse patterns in their cheeks, necks, and foreheads during a given interval.[153] However, in deepfake videos, since the video is generally layered using multiple external videos or images, there will be an inconsistency in the pulse on a subject's neck in comparison to the blood flow in their cheeks or forehead.[154]

The use of analyzing biometric data is instrumental in identifying further inconsistencies in deepfake videos. A slight difference in a subject's gait, which is their manner and style of walking,[155] or an inconspicuous difference in a person's head movements may be the difference between spotting a deepfake or a grimmer outcome: letting it plummet into the misinformation chasm that is our current internet accessible world.[156] In fact, an absence or excess of blinking has proven to be quite effective in detecting deceptive deepfakes.[157] On average, a healthy adult blinks somewhere between every two and ten seconds with a single blink taking approximately

---

MELLON UNIV. SOFTWARE ENG'G. INST. (Aug. 22, 2016), https://insights.sei.cmu.edu/blog/real-time-extraction-of-biometric-data-from-video/ [https://perma.cc/A8EL-QNHP].

[146] *Biometrics*, U.S. DEP'T. OF HOMELAND SEC., https://www.dhs.gov/biometrics [https://perma.cc/HR7E-R95J].

[147] Venneti, *supra* note 145.

[148] *Iris*, HEALTHLINE, https://www.healthline.com/human-body-maps/iris-eye#1 [https://perma.cc/9H2B-CFKR].

[149] Venneti, *supra* note 145.

[150] *See* Schmidt, *supra* note 134.

[151] Bowman, *supra* note 43.

[152] Sarah Ashley O'Brien, *Deepfakes Are Coming: Is Big Tech Ready?* CNN BUS. (Aug. 8, 2018, 11:16 AM), https://money.cnn.com/2018/08/08/technology/deepfakes-countermeasures-facebook-twitter-youtube/index.html [https://perma.cc/H7BN-B9RL].

[153] *Id.*

[154] *Id.*

[155] *Gait*, PHYSIOPEDIA, https://www.physio-pedia.com/Gait [https://perma.cc/P4BK-E9XC].

[156] Schmidt, *supra* note 134.

[157] Siwei Lyu, *The Best Defense Against Deepfake AI Might Be . . . Blinking*, FAST COMPANY (Aug. 31, 2018), https://www.fastcompany.com/90230076/the-best-defense-against-deepfakes-ai-might-be-blinking [https://perma.cc/F3J5-7MLU].

one-tenth to four-tenths of a second.[158] However, since a deepfake creating algorithm is programmed to use images of a person in which the subject's eyes are open, a deepfake algorithm is not yet sophisticated enough to create fabricated blinks, leading to potential detection by one trained to look for the anomaly.[159]

As is often apparent in technology, there is currently a deepfake war between the attackers and defenders, so detection and mitigation are absolutely necessary as the attackers become more evolved and advanced in constructing deepfakes.[160] Yet, the scientific community is constantly formulating innovative technological methods of deepfake detection including discovery of deepfake data that focuses on artifact generation.[161] In the world of computer data, an artifact is a data flaw caused by conditions, equipment, or techniques like software errors, electromagnetic interference, or miscalculations in algorithms.[162] Artifacts are then further categorized into digital, visual, compression, noise, statistical, and radar artifacts in which various flaws like distorted or corrupted images, compression issues, static noise, or ghost objects within an image result in their specific categorization.[163]

When a deepfake is created, it often generates a flaw in the image or video materializing itself as an artifact.[164] These flaws may be easily missed by humans, but can be easily detected using forensic analysis and machine learning.[165] For instance, when a deepfake video is created, the generated content is blended with the source content, creating intermingled or blurred edges in the video frame.[166] Artifacts created when spatial blending is apparent can be detected when a learning computer is taught to emphasize these artifacts within the boundaries of a video.[167] Additionally, "[t]he content of a fake face can be anomalous in context to the rest of the frame. For example, residuals from the face warping process, lighting, and varying fidelity can indicate the presence of generated content."[168]

Aside from focusing on specific artifact detection, proponents of detecting deepfakes have trained deep computer neural networks on anomaly detection models.[169] Here, a computer model is able to overcome noise and other distortions by identifying raw, original pixels in an image or

---

[158] *Id.*
[159] *Id.*
[160] Mirsky & Lee, *supra* note 31, at 1:27–28.
[161] *Id.* at 1:26.
[162] John Spacey, *7 Types of Data Artifact*, SIMPLICABLE (Apr. 16, 2017), https://simplicable.com/new/data-artifact [https://perma.cc/J4UN-BZ4E].
[163] *Id.*
[164] *Id.*
[165] Mirsky & Lee, *supra* note 31, at 1:26.
[166] *Id.*
[167] *Id.*
[168] *Id.* (internal citations omitted).
[169] *Id.* at 1:28.

video as opposed to the distorted and modified pixels of a deepfake.[170] However, just as technologically savvy and concerned individuals are using machine learning to defect deepfakes, hackers and creators of deepfake content can also evade detection via adversarial machine learning by constantly modifying and adapting the creation of deepfakes as they become aware of what detection learning methods are already being used.[171]

VIII. DAUBERT HEARINGS AND DEEPFAKES

As shown, methods for detecting deepfakes can range from rudimentary to exceedingly complex. In some instances, understanding, examining, and identifying the relevant evidence at issue pertaining to deepfakes cannot be done by a layperson. If the authenticity of a video was in question in a courtroom setting and one party was purporting that the evidence was a deepfake, a person whose occupation or means of knowledge is in a specialized field would be required to testify about the evidence. In situations similar to this, expert witnesses are often employed.[172] By definition:

> An "expert witness" is one who can enlighten a jury more than the average man in the street. A witness can be qualified as an expert if his knowledge extends beyond or supersedes that of an ordinary witness. An "expert witness" is one who is shown, either by training or experience, to be better informed than the hypothetical average juror. An expert witness, by definition, is any person whose opportunity or means of knowledge in a specialized art or science is to some degree better than that found in the average juror or witness.[173]

If a party in a lawsuit suspects deepfakes will be used in evidence, it would be the best course of action to hire and utilize an expert to combat the video or deepfake evidence or to potentially prove it is not a deepfake if adversely challenged. However, because this realm of technology is new and unchartered territory, even finding a supposed expert on deepfakes can pose a conceivable challenge.[174] As the technology continues to develop and more citizens are exposed to its potential, it is probable that experts in identifying deepfakes will materialize.

Considering the field of deepfake experts is needed and emerging, the next logical legal issue that appears when an expert witness may be used involves *Daubert* hearings. *Daubert* hearings are pre-trial hearings in which a trial court will determine the admissibility of expert testimony under the

---

[170] *Id.*

[171] *Id.*

[172] Frazier v. State, 112 So. 2d 212, 214 (Ala. Ct. App. 1958).

[173] Hinton v. State, 172 So. 3d 338, 345–46 (Ala. Crim. App. 2011) (citations omitted) (quoting Charles v. State, 350 So. 2d 730, 733 (Ala. Crim. App. 1977)).

[174] *See* Reynolds, *supra* note 74.

relevant rule of evidence along with the *Daubert* factors.[175]

In 1993, the Supreme Court announced a new test for determining the reliability of expert testimony in which the Court stressed that judges, as opposed to a closed circle of experts, will determine the reliability of expert testimony.[176] The *Daubert* decision was based on Federal Rule of Evidence 702, which governs expert testimony. Under Federal Rule of Evidence 702:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if: (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; (b) the testimony is based on sufficient facts or data; (c) the testimony is the product of reliable principles and methods; and (d) the expert has reliably applied the principles and methods to the facts of the case.[177]

Additionally, when determining if expert testimony is reliable under the rule, judges should also consider the non-exclusive *Daubert* factors which are: whether the theory or technique being implemented has been tested, whether the technique or theory has been subject to peer review and publication, the technique's error rate, the existence of standards controlling the technique's application, and whether the theory or technique has generally been accepted in the relevant scientific community.[178] Thus, if a party intends to introduce an expert witness in the dominion of deepfakes, they should meet the criteria presented by both Federal Rule of Evidence 702 and *Daubert*.

Within the evidentiary rule, 702 imposes two types of reliability including that the principles underlying the expert's approach be reliable while the application of those same principles to the facts of the case must be reliable.[179] Considering deepfake technology is already enigmatic to begin with, properly vetting deepfake experts will become a difficult, tedious, and troublesome aspect of litigation involving supposed deepfakes. For that reason, if a party to a lawsuit does attempt to challenge the authenticity of evidence on the premise that it is a deepfake, a judge will most likely hold a *Daubert* hearing based on the *Daubert* challenge:

Daubert challenges to evidence are routinely raised by a motion in *limine*, and may also require an evidentiary hearing, or "*Daubert* hearing," to properly inform the trial judge. . . .[F]ederal court[s] may decide whether a *Daubert* challenge is decided upon special briefing or some other procedure, and has further explained that a common method is a *Daubert* hearing, although such a hearing is not specifically mandated. . . . "Decisions about admissibility under all three rules [Federal Rules 403, 702, and 703] hinge on factual issues that can be resolved meaningfully only if a court is

[175] Tumlinson v. Advanced Micro Devices, Inc., 81 A.3d 1264, 1269 (Del. 2013).
[176] Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579 (1993).
[177] FED. R. EVID. 702.
[178] Elcock v. Kmart Corp., 233 F.3d 734, 745 (3d Cir. 2000).
[179] Sardis v. Overhead Door Corp., 10 F.4th 268, 281 (4th Cir. 2021).

adequately informed. . . ." A procedure creating a record of facts, whether by formal hearing with an opportunity for cross-examination, or argument on special briefs with attached documents and admitted facts of what the evidence will show, "creates a record that allows a judge to rule on admissibility after due consideration."[180]

Sometimes, the result of a *Daubert* hearing can be dispositive if a party intends to rely on a particular piece of expert testimony to support a prima facie element of their case.[181] For instance, if a party's ability to prove an element of negligence rests upon a surveillance tape, but the opposing party asserts it is a doctored deepfake video, a trial judge may rely on the testimony or opinion of a deepfake technology expert to weigh in on the authenticity of the video. Moreover, if the expert's testimony does not survive the *Daubert* hearing, the party may lack the necessary evidence needed to go to trial or if allowed to go to trial, to win the case.[182]

Since a judge's evaluation of scientific or technical evidence can be time consuming and tedious, judges will often opt for a *Daubert* hearing in advance of trial to determine the potential prejudice, reliability, and fit of the proposed evidence.[183] With that, judges may err on the side of caution with conservative decision making by being less likely to act in accordance with new scientific methods or principles. Considering deepfake technology along with deepfake detection methods are relatively new, a judge may be hesitant to include testimony relating to the authenticity or fraudulent nature of a deepfake video. This creates a double-edged sword in which a cautious or conservative judge may choose to exclude evidence that could prove a video is an altered deepfake while also potentially excluding evidence proving that a video or image is real. Since *Daubert* hearings allow judges to practice their gatekeeper function, in order to dispose of potential inequities in the courtroom, judges should become aware of the technology as well as educate themselves on the threat, use, and detection of these life altering videos.

IX. LITIGATION FOR VICTIMS AFFECTED BY DEEPFAKES

Although the major legal concern surrounding deepfakes is the potential introduction of deepfake evidence into the courtroom as authenticated evidence, the legal realm will also be affected by victims of deepfake technology seeking legal recourse. In the future, there could be potential for ample litigation by victims of deepfakes relying on various tort

[180] Christian v. Gray, 65 P.3d 591, 610 (Okla. 2003) (internal citations omitted).
[181] *See* MERRITT & SIMMONS, *supra* note 77, at 804–05.
[182] *Id.*
[183] *Id.*

and fraud legal theories.[184] In common law jurisdictions, the victim of a deepfake, in theory, can sue the deepfake's creator under a privacy tort.[185] The most applicable tort is probably the "false light" theory. A false light action permits recovery for injuries "caused by publicity that unreasonably places the plaintiff in a false light before the public."[186] A claim of this nature "place[s] a person in a false light even though the facts themselves may not be defamatory."[187] As one of the four recognized claims in the area of invasion of privacy, a person can be liable under this doctrine if they publicize or share a matter involving another person in a false light where it would be highly offensive to the reasonable person.[188] Here, "'the essence of the claim' is the falsity of what the publication communicates."[189]

Additionally, if a deepfake is being used to improperly promote a product or service, the victim or subject of the deepfake may be able to invoke the misappropriation doctrine.[190] Also a privacy tort, misappropriation could help the victim recover any profits made from the commercial use of their image or likeness in addition to any applicable statutory or punitive damages.[191] This doctrine can even be applied in tandem with a false light claim where applicable.[192]

Similarly, if the deepfake content asserts factually untrue proclamations, traditional causes of action, like libel or defamation, may lead to a victim's victory in court. A victim of deepfake technology may also employ the legal weapon of intentional infliction of emotional distress, which stems from defamation or libel. Since deepfakes often portray a victim in an unusual or erroneous light, a person may be able to meet the elements of an intentional infliction of emotional distress if they can show: (1) intentional or reckless conduct; (2) extreme or outrageous conduct; (3) a causal connection between the wrongful conduct and the emotional distress; and (4) severe emotional distress.[193] These elements are tough to prove, specifically if the deepfake employs parodical or satirical use of a person's image.[194] Under fair use and parody law, if a deepfake imitates a serious piece of work as a method of criticism and meets the elements of the doctrine, a deepfake creator may escape liability under the fair use

---

[184] Riana Pfefferkorn, *'Deepfakes': A New Challenge for Trial Courts*, NW SIDEBAR (Mar. 13, 2019), https://nwsidebar.wsba.org/2019/03/13/deepfakes-a-new-challenge-for-trial-courts/ [https://perma.cc/S88T-JZMU].

[185] Schmidt, *supra* note 134.

[186] Cain v. Hearst Corp., 878 S.W.2d 577, 580 (Tex. 1994).

[187] Straub v. Lehtinen, Vargas & Riedi, P.A., 980 So. 2d 1085, 1086 (Fla. Dist. Ct. App. 2007).

[188] *Id.*

[189] *Id.* at 1086–87.

[190] Schmidt, *supra* note 134.

[191] *Id.*

[192] *Id.*

[193] Davenport v. Maryland, 38 F. Supp. 3d 679, 684 (D. Md. 2014).

[194] *See, e.g.*, Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

section of the Copyright Act.[195]

Aside from tortious claims, victims of deepfakes may also be able to prosecute under criminal laws. Depending on the inception, purpose, and nature of the deepfake, the subject of a deepfake could potentially bring a claim under domestic abuse statutes, sexual harassment statutes, cyberbullying laws, or even cyberstalking.[196] One remedy of law that may be of particular importance to the subject of a deepfake video is an injunction against the deepfake itself, which would attempt to remove the deepfake from the internet altogether and optimistically limit the unintended effects and reach of the video.[197]

X. A TECHNOLOGICAL CALL TO ACTION

Deepfakes, like all technology, will become rapidly and more readily available to users as the technology continues to advance. As deepfake detection methods become more commonplace and obtainable, deepfake originators will be consistently innovating new ways to evade detection and continue their use of these altered videos, photos, and audio. As a society, public figures in the education, political, entertainment, and specifically legal field need to be aware of the risk and threat of deepfakes while presenting the issues to their respective constituents.

In the realm of social media, major players in the field have announced policies, although limited, relating to monitoring and policing deepfakes.[198] Facebook said it would remove "manipulated misleading media" which has been "edited or synthesized" using machine learning if the videos mislead someone into thinking the subject of the video did or said something they did not actually partake in.[199] Likewise, Reddit, the major message board website, said it will remove media that "impersonates individuals or entities in a misleading or deceptive manner."[200] While these policies will assist in removing glaring and obvious depictions of deepfakes, current social media policies leave room for deepfakes to slide through the cracks and infiltrate millions of users' news feeds.[201] To combat this, education and awareness of so-called "fake news" perpetrated by deepfake technology needs to become widespread.

---

[195] *Parody: Fair Use or Copyright Infringement*, FINDLAW (June 8, 2017), https://corporate.findlaw.com/intellectual-property/parody-fair-use-or-copyright-infringement.html [https://perma.cc/RW8A-F64A].

[196] Schmidt, *supra* note 134.

[197] *Id.*

[198] James Vincent, *Facebook's Problems Moderating Deepfakes Will Only Get Worse in 2020*, THE VERGE (Jan. 15, 2020), https://www.theverge.com/2020/1/15/21067220/deepfake-moderation-apps-tools-2020-facebook-reddit-social-media [https://perma.cc/7TYN-ABK2].

[199] *Id.*

[200] *Id.*

[201] *Id.*

Moving to the courtroom, as education and awareness of deepfakes grows, the effects on the legal system will be varied. For judges and lawyers alike, advocates within the legal system need to be armed with the knowledge of what exactly a deepfake is along with a cautioned and careful analysis of video and photographic evidence entered into evidence. Photo and video analysis experts may emerge as an integral and necessary aspect of cases involving video or photographic evidence as the technology continues to advance. Judges may call for pre-trial *Daubert* hearings to determine the reliability of expert evidence about deepfakes before the jury ever has a chance to view the potentially damning evidence.

Similarly, deepfakes may rear their ugly head as jurors deliberate questions of fact surrounding photos and videos. A well-aware and tech savvy juror may question every video and photograph entered into evidence with extreme inspection while the unaware and potentially naïve juror may take videos at face value by giving the video evidence heartier weight than should be allotted. Legal advocates will need to take into consideration the values, understanding, and overall knowledge jurors have about deepfake technology if a case involves audio, video, or photographic evidence.

XI. CONCLUSION

Deepfake technology is an ever-evolving realm of digital modification that allows users to create videos, audio, or photos of a person partaking in an activity that person has never done and would never do. Deepfakes are becoming rampant in the worlds of social media, entertainment, and politics. A young but considerable danger of deepfake technology is making its implication in the legal system. The history of evidence and authentication of evidence has led to a lenient standard in which currently, most videos and photos are admissible unless there is just cause to believe otherwise. In the realm of the courtroom, deepfakes cause concern relating to evidentiary admission as well as the style and method attorneys may use in litigation to zealously advocate for their clients. As detection methods of deepfakes grow more widespread, creators of deepfakes will constantly be at war with those combatting it, trying always to be multiple steps ahead and ensure their deepfake content will prevail. Legal scholars, social media platforms, and news stations alike need to alert the general public to the use of deepfakes so the common internet user will question the images presented to them and, in turn, prevent the creation of an alternate reality.