The Deepfake Defense—Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers

REBECCA A. DELFINO*

"Faith is to believe what you do not see; the reward of this faith is to see what you believe." — Saint Augustine

TABLE OF CONTENTS

* Associate Dean for Clinical Programs and Experiential Learning, and Professor of Law at LMU Loyola Law School, Los Angeles. I would like to thank Matthew Babb, Lindsey Khim, Erica Barseghian, Blake Ackerman and Melis Tirhi their research assistance, Loyola Law School for its support of faculty scholarship, and as always, my family for their patience, indulgence, and encouragement. This article is dedicated to Melissa Sexton.

I. INTRODUCTION

Thousands of audiovisual images documented the insurrectionists who stormed the United States Capitol on January 6, 2021.[1] Journalists, innocent bystanders, members of congress, and law enforcement captured images that they widely disseminated in live-stream videos and on the internet.[2] Even some insurrectionists created videos and social media posts boasting of their

[1] *See* Olivia Rubin, Alexander Mallin & Will Steakin, *By the Numbers: How the Jan. 6 Investigation Is Shaping up 1 Year Later*, ABC NEWS (Jan. 4, 2022), https://abcnews.go.com/US/numbers-jan-investigation-shaping-year/story?id=82057743 [https://perma.cc/CA8Y-PZS2].

[2] *See* Julio Cortez & Andrew Harnik, *Images of Chaos: AP Photographers Capture US Capital Riots*, AP NEWS (Jan. 5, 2022), https://apnews.com/article/photos-election-jan6-trump-Washington-f69b5f03316eaef2044d520bc7ffe49a [https://perma.cc/W48Q-P3VU]; Kat Lonsdorf, Courtney Dorning, Amy Isackson, Mary Louise Kelly & Alisa Chang, *A Timeline of How the Jan. 6 Attack Unfolded—Including Who Said What and When*, NPR (June 9, 2022), https://www.npr.org/2022/01/05/1069977469/a-timeline-of-how-the-jan-6-attack-unfolded-including-who-said-what-and-when [https://perma.cc/DB3W-4E7H].

involvement.[3] The FBI has subsequently collected more than 23,000 video files recording the events.[4] Based on the audiovisual evidence, many insurrectionists have been arrested; thus far, more than 750 people have been charged for their role in the insurrection, and the investigation continues to pursue hundreds more.[5] Given the overwhelming audio and visual evidence implicating some insurrectionists, it should be impossible to assert a plausible defense based on the claim that those clearly depicted in the images were not present. Right? Wrong. The emergence of deepfakes has changed the landscape of plausible defenses.

As the defense in the criminal trial of January 6th insurrectionist leader Guy Reffitt illustrates, in the age of deepfakes, it is easier than ever for lawyers to exploit juror suspicions about what is real and what is fake. Reffitt, an alleged member of the anti-government movement, the Three Percenters, drove from Texas to attend the January 6th pro-Trump rally in Washington D.C.[6] Dressed in riot gear and carrying a firearm, Reffitt stood at the front of the crowd with a megaphone, steering the attack on the Capitol.[7] Videos and other visual images showed him leading the crowd advancing toward a police line of defense above the Capitol's West Terrace that afternoon.[8] Reffitt was later arrested and charged with multiple crimes, including bringing a gun to the insurrection, obstruction, and assaulting the police.[9] And although video, audio, text messages, the testimony of his son, and another insurrectionist clearly and unmistakably implicated Reffitt in the crimes, his lawyer attempted to undermine the evidence, suggesting the incriminating images may have been fabricated.[10] His lawyer told the jury that the evidence against Reffitt was a "deepfake"[11]—an audiovisual recording created using readily available Artificial Intelligence (AI) technology that allows anyone with a smartphone to

---

[3] *See* Sandra Jeppesen, *The January 6 Insurrection Showed that Performance Crime Is Becoming Increasingly Popular*, CONVERSATION (July 7, 2022), https://theconversation.com/the-jan-6-insurrection-showed-that-performance-crime-is-becoming-increasingly-popular-186102 [https://perma.cc/5J3J-C4X4] (describing the January 6th insurrectionists' efforts to document their actions on video, text and other media).

[4] Rubin, Mallin & Steakin, *supra* note 1.

[5] *Id.*

[6] *See* Dana Verkouteren, *In the First Jan. 6 Trial, a Jury Found Capitol Riot Defendant Guy Reffitt Guilty*, DIGIS MAK (Mar. 8, 2022), https://digismak.com/in-the-first-jan-6-trial-a-jury-found-capitol-riot-defendant-guy-reffitt-guilty/ [https://perma.cc/YRL3-GGR5].

[7] *Id.*; Rachel Weiner & Spencer S. Hsu, *Texas Man Guilty on All Courts in First Capitol Riot Trial*, TEX. TRIB. (Mar. 8, 2022) https://www.texastribune.org/2022/03/08/capitol-riot-trial-guy-reffitt/ [https://perma.cc/BV7N-4D37].

[8] Alan Feuer, *Texas Man Convicted in First Jan. 6 Trial*, N.Y. TIMES (Mar. 8, 2022), https://www.nytimes.com/2022/03/08/us/politics/guy-reffitt-jan-6-trial.html [https://perma.cc/NK6G-Q82G]; Weiner & Hsu, *supra* note 7.

[9] *Id.*

[10] *See* Verkouteren, *supra* note 6.

[11] *Id.*

believably map one person's movements and words onto another person.[12] Current law does not provide a clear response to Reffitt's lawyer's reliance on deepfakes as a defense.

But this is clear—the deepfake defense is a new danger to our legal system's adversarial process and truth-seeking function. The "deepfake defense" is built around the premise that the audiovisual material introduced as evidence against the defendant is claimed to be fake. The existence of deepfakes presents challenges in court proceedings because they create an opportunity to raise objections and arguments to even genuine evidence.[13] Moreover, because the norms of professional ethics expect lawyers to advocate zealously[14], the fact that deepfakes exist invites lawyers to exploit juror bias and skepticism about what is real. Thus, lawyers may plant the seeds of doubt in jurors' minds to question the authenticity of all audiovisual images even where the lawyers know the evidence is real.[15]

Currently, no rule of procedure, ethics, or legal precedent directly addresses the presentation of the "deepfake defense" in court.[16] The existing legal and ethical standards provide scant guidance because they were developed before the advent of deepfake technology.[17] As a result, they do not deter lawyers from exploiting their existence.[18] Although legal scholarship and the popular news media have addressed certain facets of deepfakes, there has been no in-depth commentary on the deepfake defense.

This article addresses matters that prior academic scholarship has not yet examined. It will explore the contours of the deepfake defense, locating it within the historical and current framework of lawyers' efforts to fabricate and manipulate evidence and challenge the authenticity of the evidence. The article will also consider the laws and the practice norms to curb that misconduct. This examination implicates the model rules of professional conduct, the rules of procedure, and the substantive law. It will also propose reconsidering the ethical rules governing candor, fairness, and the limits of zealous advocacy and urge a re-examination of the court's role in sanctioning conduct under the rules of procedure. Thus, this article will start that conversation and offer proposals to guide the way forward for lawyers and courts as they traverse this new technological landscape.

---

[12] *See* Rebecca A. Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 FORDHAM L. REV. 887, 892–94, 929 (2019) [hereinafter Delfino, *Pornographic Deepfakes*].

[13] Rebecca A. Delfino, *Deepfakes on Trial: A Call to Expand the Trial Judge's Gatekeeping Role to Protect Legal Proceedings from Technological Fakery*, 74 HASTINGS L.J. 293, 305–06 (2023) [hereinafter Delfino, *Deepfakes on Trial*].

[14] MODEL RULES OF PRO. CONDUCT r. 1.3 cmt. 1 (AM. BAR ASS'N 2021).

[15] *See, e.g.*, Verkouteren, *supra* note 6.

[16] Delfino, *Deepfakes on Trial*, *supra* note 13, at 332.

[17] *Id.* at 306–07.

[18] *Cf.* MODEL RULES OF PRO. CONDUCT r. 3.3 (AM. BAR ASS'N 2021).

Part II defines and explains the rise of deepfakes and the phenomenon known as the "Liar's Dividend," which invites challenges to the authenticity of genuine digital audio-visual images by suggesting that they are too deepfakes.[19] This Part explores the emergence of Liar's Dividend in legal proceedings through lawyers' use of the "deepfake defense."

It further considers the unique dangers the deepfake defense poses to the justice system and the jury's fact-finding process. It addresses the growing distrust and doubt among potential jurors about the authenticity of all audiovisual evidence. This Part also argues that a lawyer's reliance on the deepfake defense to mislead and distract the jury falls beyond the bounds of acceptable lawyer conduct.

Part III describes the current legal and ethical mechanisms to address the use of the deepfake defense. It begins by discussing how legal systems have historically dealt with lawyers who engage in deception and what can be learned from that history. This Part also explores the complexity of the legal and prudential issues implicated by deepfake defense in legal proceedings by examining the rules of civil procedure, substantive law, and ethical rules that govern lawyers' trial conduct. Part III also identifies the shortcomings and limitations of the current mechanisms available to deter the use of the deepfake defense. It argues that none of the current mechanisms can fully address the challenges posed by deepfakes in the courtroom, including problems with corruption of the trial process and juror bias and skepticism.

Finally, Part IV offers solutions to the presentation of the deepfake defense in legal proceedings grounded revisions to the rules governing the conduct of lawyers and trial proceedings. The Article argues that addressing the deepfake defense in legal proceedings will require changes to the rules of procedure, substantive law, and professional conduct. Specifically, an amendment to Federal Rule of Civil Procedure Rule 11 to broaden its reach to apply to counsel's oral representation is urged. This article also recommends clarification of the law governing unreasonable and vexatious trial conduct to cover the use of the deepfake defense. Finally, a revision to the model rules of ethics would minimize jury skepticism and bias and guard against the corruption of the judicial system's truth-determining process that results from the use of the deepfake defense.[20]

---

[19] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1758 (2019).

[20] In addition, to the challenges imposed by the deepfake defense on the conduct of trials, functioning of the jury and behavior of lawyers, the presentation of deepfakes in legal proceedings also raises questions regarding the admission of evidence. However, the implications on the rules of evidence are beyond the scope of this article but are explored elsewhere in the legal scholarship. *See, e.g.*, Delfino, *Deepfakes on Trial*, *supra* note 13, at 340–41.

II. DEEPFAKES, THE LIAR'S DIVIDEND AND THE DEEPFAKE DEFENSE—SCOPE AND CONTEXT

The critique of the legal rules and ethical norms implicated by a lawyer's use of the deepfake defense in court begins with a brief background on the deepfake phenomenon and the dangers they have posed outside of the legal of the courtroom.

A. *The Rise of Deepfakes and the "Liar's Dividend"*

Deepfakes are fabricated audiovisual content created or altered to appear to the observer to be a genuine account of the speech, conduct, image, or likeness of an individual or an event.[21] They create a fake reality by superimposing a person's face on another's body or changing the contents of one's speech.[22] A combination of "deep learning" and "fake," so-called "deepfake" programs use AI to produce these fake audio-visual images.[23] This new technology, developed and unleashed on the internet in late 2017, allows anyone with a smartphone to believably map another's movements and words onto someone else's face and voice to make them appear to say or do anything.[24] And the more video and audio of the person fed into the computer's deep-learning algorithms, the more convincing the result.[25]

As scholars and technologists predicted, deepfakes have proliferated.[26] They have been used as entertainment[27] and weapons to exploit and traumatize—often women.[28] The use of deepfakes has also spread into the

---

[21] Delfino, *Pornographic Deepfakes*, *supra* note 12, at 892–94, 929.

[22] *Id.* at 929; Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 339–40 (2019) (referencing deepfake technology's pornographic beginnings).

[23] Delfino, *Pornographic Deepfakes*, *supra* note 12, at 889, 892–94, 929. The term "deepfake" is derived from a combination of the phrases "deep learning" and "fake." *Id.* at 292; Douglas Harris, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 99–100 (2019) (quoting Sundar Pichai of Google whose company invented TensorFlow to develop artificial intelligence tools for the public).

[24] *See* Delfino, *Pornographic Deepfakes*, *supra* note 12, at 887; Spivak, *supra* note 22, at 339 (referencing deepfake technology's pornographic beginnings).

[25] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 299. Deep learning refers to a training process by which the AI technology becomes increasingly intelligent through the continued introduction of information into the system. *Id.* Deepfake software applications work by uploading digital images into a "machine-learning algorithm that's trained itself to stitch one face on top of another." Delfino, *Pornographic Deepfakes*, *supra* note 12, at 893.

[26] *See generally* Chesney & Citron, *supra* note 19, at 1769–77 (describing the expanse of domestic, social, economic and foreign policy implications of deepfakes).

[27] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 300 (discussing examples of deepfakes used in entertainment).

[28] *See id.* at 300–01. More than 90% of deepfakes on the internet are pornographic depictions of women. HENRY AJDER, GIORGIO PATRINI, FRANCESCO CAVALLI & LAURENCE

political sphere to create fake news and false images involving political leaders and governmental actors.[29]

Deepfakes pose dual dangers.[30] The obvious threat stems from the potential that the disinformation in the *fake* image will win over its audience to convince the viewer that something fake is real. This danger has garnered significant attention from legal scholars, multiple journalists, and technologists.[31]

CULLEN, THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT 1–2 (Sept. 2019), https://regmedia.co.uk/2019/10/08/deepfake_report.pdf [https://perma.cc/LGD6-S3D4]. Some subjects are celebrities, but private citizens have been exploited as well. Cleo Abram, *The Most Urgent Threat of Deepfakes Isn't Politics. It's Porn.*, VOX (June 8, 2020), https://www.vox.com/2020/6/8/21284005/urgent-threat-deepfakes-politics-porn-kristen-bell [https://perma.cc/3855-N8W5].

[29] *See* Dave Lee, *Deepfakes Porn Has Serious Consequences*, BBC NEWS (Feb. 3, 2018), https://www.bbc.com/news/technology-42912529 [https://perma.cc/R4TY-LJQK] (referencing a claim as early as 2018 that deepfake technology "could down the road be used maliciously to hoax governments and populations or cause international conflict"). For instance, in 2019, millions of people viewed a video with altered audio showed of United Stated House of Representative Speaker Nancy Pelosi appearing to slur her speech while speaking during a recorded interview. 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]. The video was circulated widely on social media, including a Tweet by former President Donald Trump. *Id.*; *see* Lauren Feiner, *Facebook Says the Doctored Nancy Pelosi Video Used to Question Her Mental State and Viewed Millions of Times Will Stay Up*, CNBC (May 24, 2019), https://www.cnbc.com/2019/05/24/fake-nancy-pelosi-video-remains-on-facebook-and-twitter.html [https://perma.cc/ZGZ2-G6S4] (emphasizing the impact of the deepfake video of Nancy Pelosi on public perception). In Malaysia, a politician was mired in a sex tape scandal after a deepfake video surfaced purporting to show him engaging in illegal homosexual activity. Jarni Blakkarly, *A Gay Sex Tape Is Threatening to End the Political Careers of Two Men in Malaysia*, SBS NEWS (June 17, 2019), 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]; *see also* Digital Forensic Research Lab, *Russian War Report: Hacked News Program and Deepfake Video Spread False Zelenskyy Claims*, ATL. COUNCIL (Mar. 16, 2022) https://www.atlanticcouncil.org/blogs/new-atlanticist/russian-war-report-hacked-news-program-and-deepfake-video-spread-false-zelenskyy-claims/ [https://perma.cc/KMX6-4WD8] (describing the use of deepfake videos in the Ukraine and Russian war); Jane Wakefield, *Deepfake Presidents Used in Russia-Ukraine War*, BBC NEWS (Mar. 18, 2022), https://www.bbc.com/news/technology-60780142 [https://perma.cc/J8ZU-VWF9].

[30] *See* Matteo Wong, *We Haven't Seen the Worst of Fake News*, ATLANTIC (Dec. 22, 2022), https://www.theatlantic.com/technology/archive/2022/12/deepfake-synthetic-media-technology-rise-disinformation/672519/ [https://perma.cc/66HP-V56V] (summarizing the ongoing threat of deepfakes to political and cultural discourse and institutions, including the increasing exposure of the public to deepfake technology and AI technologies).

[31] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 302–05 (citing sources); *see also* Rob Toews, *Deepfakes Are Going to Wreak Havoc on Society We Are Not Prepared*, FORBES (May 25, 2020), https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=70136e487494 [https://perma.cc/WPQ4-CF96]; Karen Hao, *The Biggest Threat of Deepfakes Isn't the Deepfakes Themselves*, MIT

This article focuses on the other less apparent but equally dangerous threat that has arisen in the deepfake era, which has received scant attention in the scholarship on deepfakes—the knowledge that deepfakes exist will cause viewers to believe that a genuine image is fake. This is known as the "Liars Dividend"—a term coined by Professors Robert Chesney and Danielle Keats Citron to describe the phenomenon that arises when the ability to create convincing fakes allows the image creators to undermine the veracity of *real* information by claiming that it, too, is a fabrication.[32] Lies not only involve facts that never happened, but falsehoods are also very often disseminated to undermine the truth. Deepfakes make it easier for the liar to deny the truth.[33] An accusation may be supported by real video or audio evidence. But as the public becomes more aware that video and audio can be convincingly faked, liars will exploit this awareness to escape accountability for their actions by denouncing authentic video and audio as deepfakes.[34] Even technology experts fear that a skeptical public will be primed to doubt the authenticity of real audio and video evidence.[35]

This predicted harm is not theoretical. As an example of the Liar's Dividend at work, in 2017, President Ali Bongo Ondimba of Gabon suffered a stroke had spent several months out of the country, seeking medical treatment.[36] As a result, rumors circulated that the President was dead. To alleviate public concern and quell the rumors that he was dead or incapacitated, on New Year's Day 2018, the president released a "proof of life" video to show that he was recovering from a stroke.[37] However, the video was attacked, fueling speculation about Bongo's condition.[38] Days later, the president's opponents claimed that this video was a deepfake and attempted a coup.[39]

TECH. REV. (Oct. 10, 2019), https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/ [https://perma.cc/4SNY-8D33].

[32] Chesney & Citron, *supra* note 19, at 1758.

[33] *Id.*

[34] *See id.*

[35] Hany Farid, an AI and deepfakes researcher and associate dean of UC Berkeley's School of Information, explained that his "biggest concern is not the abuse of deepfakes, but the implication of entering a world where any image, video, audio can be manipulated. In this world, if anything can be fake, then nothing has to be real, and anyone can conveniently dismiss inconvenient facts" as fake images. Tyler Sonnemaker, *'Liar's Dividend': The More We Learn About Deepfakes, the More Dangerous They Become*, BUS. INSIDER (Apr. 13, 2021), https://www.businessinsider.com/deepfakes-liars-dividend-explained-future-misinformation-social-media-fake-news-2021-4 [https://perma.cc/V6TJ-BJJU].

[36] *See* Sarah Cahlan, *How Misinformation Helped Spark an Attempted Coup in Gabon*, WASH. POST (Feb. 13, 2020), https://www.washingtonpost.com/politics/2020/02/13/how-sick-president-suspect-video-helped-sparked-an-attempted-coup-gabon/ [https://perma.cc/4S5L-E9GL].

[37] *Id.*

[38] *Id.*

[39] *Id.*

Closer to home, another attempt to exploit the Liar's Dividend, though ultimately unsuccessful, happened after the video of George Floyd's death went viral. Two weeks after Floyd's death, a Republican congressional candidate from Missouri posted a 23-page "report" pushing a conspiracy theory that Floyd had died years earlier and that someone had used deepfake technology to superimpose his face onto the body of an ex-NBA player to create a video to stir up racial tensions.[40]

And as in the case of deepfakes, the impact of the Liar's Dividend is not confined to the realm of politics, public figures, or newsworthy events. Private individuals have also had their lives disrupted and destroyed by the deployment of the Liar's Dividend to deny the truth. For instance, in the spring of 2021, a Pennsylvania woman, Raffaela Spone, was arrested and charged with multiple counts of harassment for allegedly creating deepfakes to frame her daughter's cheerleading rivals.[41] The prosecutor claimed that Spone—who became known in the media as "deepfake cheerleader mom" created a fake video of the teenage girl vaping, that she altered social media accounts of the girls to make them appear nude, drinking and vaping, and that she sent them texts and voicemails telling them to kill themselves.[42] One of the victims even appeared on ABC's Good Morning America to claim she had been the subject of Spone's deepfake.[43] Spone denied creating the deepfakes; ultimately, a firm of technology experts volunteered to help her.[44] Those digital-forensic experts determined that the alleged fake videos were authentic.[45] Thus, the alleged victims employed the Liar's Dividend to escape responsibility for their conduct depicted in the videos. They used the Liar's Dividend to dupe the prosecutor into filing a criminal complaint against Spone, and ultimately, the prosecutor

---

[40] Will Sommer, *GOP House Candidate Insists George Floyd Killing Was Staged*, DAILY BEAST (June 24, 2020), https://www.thedailybeast.com/gop-house-candidate-winnie-heartstrong-insists-george-floyd-killing-was-staged [https://perma.cc/LS6E-5PQK].

[41] Kim Bellware, *Cheer Mom Used Deepfake Nudes and Threats to Harass Daughter's Teammates, Police Say*, WASH. POST (Mar. 13, 2021), https://www.washingtonpost.com/nation/2021/03/13/cheer-mom-deepfake-teammates/ [https://perma.cc/R422-YJVA].

[42] *See id.*; *Cheerleader's Mom Accused of Making "Deepfake" Videos of Daughter's Rivals*, CBS NEWS (Mar. 15, 2021), https://www.cbsnews.com/news/raffaela-spone-cheerleader-mom-deepfakes/ [https://perma.cc/B36G-XBDF]; *Mother 'Used Deepfake to Frame Cheerleading Rivals*,' BBC NEWS (Mar. 15, 2021), https://www.bbc.com/news/technology-56404038 [https://perma.cc/N7DQ-BC44]; E-mail from Robert J. Birch, Esq., Law Offices of Robert J. Birch, to author (Feb. 2, 2022) (on file with author) [hereinafter Birch E-mail].

[43] Drew Harwell, *Remember the 'Deepfake Cheerleader Mom'? Prosecutors Now Admit They Can't Prove Fake-Video Claims*, WASH. POST (May 14, 2021), https://www.washingtonpost.com/technology/2021/05/14/deepfake-cheer-mom-claims-dropped/ [https://perma.cc/8VZG-3YED].

[44] Birch E-mail, *supra* note 42.

[45] *Id.*; Harwell, *supra* note 43.

decided he could no longer use the video as a basis for the charges.[46] But by the time the prosecution had changed course, the damage to Spone had been done. According to her lawyer, her reputation has been destroyed; she received death threats and was ridiculed and harassed in her community.[47]

B. *The Liar's Dividend in Legal Proceedings Is "The Deepfake Defense"*

Today, video and audio recordings are indispensable elements of many criminal and civil actions.[48] As the case of deepfake cheerleader mom illustrates, deepfakes and, more specifically, the Liar's Dividend, have invaded the legal system. Even though our legal system is as vulnerable to content manipulation as any other area of civic life, legislative efforts have not explicitly addressed the impact of deepfakes on the broader justice system.[49]

At its essence, the common law adversarial system depends upon the parties offering competing versions of the claims through the presentation of evidence. Before a court permits evidence to be presented to the trier of fact—the jury——the proponent is required to prove the evidence is real; in other words, it is an authentic representation of what it purports to show.[50] Lawyers, courts, and

---

[46] Birch E-mail, *supra* note 42; Harwell, *supra* note 43. At a subsequent hearing in July of 2021, the detective on the case testified that they do not have evidence that the video was deepfaked, that there is no evidence that Spone manipulated any social media images, and that there were no threats. *Id.*

[47] Birch E-mail, *supra* note 42; Harwell, *supra* note 43.

[48] Reliance on audiovisual evidence at all stages of legal proceedings from the investigation stage through the trial phase is increasing and important. *See, e.g.*, NAT'L FORENSIC SCI. TECH. CTR., A SIMPLIFIED GUIDE TO FORENSIC AUDIO AND VIDEO ANALYSIS 2 (discussing that the use of audiovisual evidence in the investigation stage has increased exponentially); Snowden Becker & Jean-François Blanchette, *On the Record, All the Time: Audiovisual Evidence Management in the 21st Century*, D-LIB MAG. (2017), https://doi.org/10.1045/may2017-becker [https://perma.cc/8AYK-9V77] ("We are currently witnessing the explosive growth of audiovisual evidence generated by widespread deployment of surveillance cameras, smartphones, and bodycams in law enforcement."); Mark Kersten, *Challenges and Opportunities: Audio-Visual Evidence in International Criminal Proceedings* (Mar. 4, 2020), https://justiceinconflict.org/2020/03/04/challenges-and-opportunities-audio-visual-evidence-in-international-criminal-proceedings/ (examining how audiovisual evidence has a history of being and still is a valuable source of evidence in international criminal courts and tribunals (ICTs)); Emily Riley, *Putting Your Smart Phone on the Witness Stand*, CRIME REP. (June 3, 2021), https://thecrimereport.org/2021/06/03/putting-your-smart-phone-on-the-witness-stand/ [https://perma.cc/94K9-5UNW] (addressing the proliferation of audio-visual evidence in trials).

[49] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 302–05 (summarizing the state of state and federal legal responses to deepfakes).

[50] *See* FED. R. EVID. 901(a).

juries have considered the evidence this way for hundreds of years.[51] This process of determining the truth has effectively functioned because evidence could be quickly, efficiently, and reliably evaluated. The system has relied on the effectiveness of the innate human ability to determine what is real by trusting one's senses under the theory—"seeing is believing."[52] Deepfake evidence and the Liar's Dividend upend this process. As deepfake technology continually improves and inches closer to becoming indistinguishable from reality, judges and jurors will be hard-pressed to determine from the naked eye whether the evidence represents the truth of the matter asserted.[53] The means of detecting deepfakes have not kept pace with the technology used to create them.[54] And thus, the introduction of deepfake evidence and the use of the Liar's Dividend in the courtroom raises new and profound issues for the administration of justice in civil and criminal proceedings.[55]

Deepfakes evidence may arise in the legal proceedings in two circumstances—as the subject of a crime or civil claim or as item evidence offered in case to prove an unrelated claim.[56] The Liar's Dividend may appear in any legal proceeding where digital and audio images are presented as a part of the proof in the case.

Deepfakes present several challenges for counsel and the court centering on proving digital image or audio evidence is authentic (or not). Deepfakes as proof

---

[51] Delfino, *Deepfakes on Trial*, *supra* note 13, at 307; *see* Paul W. Grimm, Daniel J. Capra & Gregory P. Joseph, *Authenticating Digital Evidence*, 69 BAYLOR L. REV. 1, 4 n.9 (2017) (noting the history of authenticating traditional forms of electronic evidence).

[52] *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].

[53] *See generally* Agnieszka McPeak, *The Threat of Deepfakes in Litigation: Raising the Authentication Bar to Combat Falsehood*, 23 VAND. J. ENT. & TECH. L. 433 (2021) (commenting on the speed at which deepfakes are developing and the corresponding responsibility attorneys have to challenge any evidence that purports itself as fake).

[54] *See* Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J. L. & TECH. 1, 25–26 (2020) (discussing the belief among the community of computer science and digital forensics experts that detection methods cannot keep pace with the innovations aimed at evading detection).

[55] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, 305–13, 321–48 (describing the challenges in presenting audio-digital evidence in the age of deepfake technology and proposing solutions, including changes to the Federal Rules of Evidence); Riana Pfefferkorn, *"Deepfakes" in the Courtroom*, 29 B.U. PUB. INT. L.J. 245, 254 (2020); *see also* Kathryn Lehman, Scott Edson & Victoria Smith, *5 Ways to Confront Potential Deepfake Evidence in Court*, LAW 360 (July 26, 2023), https://www.law360.com/articles/1181306/5-ways-to-confront-potential-deepfake-evidence-in-court [https://perma.cc/NTQ8-66TU] (explaining what Deepfake Technology is and how distinct it is from general manipulated audio and video).

[56] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 305–13.

of a case have already arisen in defamation cases[57] and a federal civil rights action.[58] These challenges will require revision to the rules of evidence.[59]

The second distinct challenge in legal proceedings—the central focus of this article—concerns a lawyer's use of the Liar's Dividend in court proceedings. In this courtroom context, the "Liar's Dividend" manifests as a "deepfake defense." Just as the Liar's Dividend fueled the attempted coup in Gabon, permitted the congressional candidate to exploit the death of George Floyd, and encouraged the cheerleader to accuse a rival's mother of criminal harassment,[60] the deepfake defense allows lawyers, such as Guy Reffitt's trial lawyer, to raise questions, objections, and arguments to challenge even genuine evidence. The fact that deepfakes exist invites parties (and their lawyers) to exploit their existence—to plant the seeds of doubt in jurors' minds to question the authenticity of all digital audio and images, even where the lawyer knows the evidence is authentic.

The deepfake defense is not a remote and theoretical threat to the administration of justice and court proceedings. The deepfake defense arises in two contexts. First, the deepfake defense arises during the oral argument, when counsel urges the jury to doubt the veracity of visual evidence. Counsel offers this argument solely to exploit the growing distrust and bias among potential jurors about the authenticity of all digital and audio image evidence. Psychological research reveals that humans value visual perception above other indicators of truth.[61] Thus, humans often accept images and other digital media at face value.[62] The legal system has historically shown a strong favorability

---

[57] *See In re* Woori Bank, No. 21-mc-80084-DMR, slip op. at 1–2 (N.D. Cal. June 6, 2021) (plaintiff sought discovery from social media platform to support his defamation action based on claim that a "deepfake" image of the plaintiff engaging in an improper intimate act had been posted on a social media platform).

[58] *See* Hohsfield v. Staffieri, No. 21-19295 (FLW), slip op. at 1 (D.N.J. Nov. 1, 2021) (plaintiff brought a 42 U.S.C. § 1983 action against police officers, claiming that they created a deepfake photo of him engaging in a lewd act to frame him and justify his arrest).

[59] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 305–13, 321–48 (describing the challenges in presenting audio-digital evidence in the age of deepfake technology and proposing solutions, including changes to the Federal Rules of Evidence).

[60] *See supra* notes 37–43 and accompanying text.

[61] *See* Purnell, *supra* note 52.

[62] *See* Richard K. Sherwin, Neal Feigenson & Christina Spiesel, *Law in the Digital Age: How Visual Communication Technologies Are Transforming the Practice, Theory, and Teaching of Law*, 12 B.U. J. SCI. & TECH. L. 227, 246 (2006). Studies have shown over and over again that people tend to believe what they see, despite knowing that videos can misrepresent facts. *See* Yael Granot, Emily Balcetis, Neal Feigenson & Tom Tyler, *In the Eyes of the Law: Perception Versus Reality in Appraisals of Video Evidence*, 24 PSYCH. PUB. POL'Y & L. 93, 97–98 (2018) (warning how powerful video evidence can be in convincing people that a fake event occurred). In a study conducted by a bank where no participants illicitly took money, the bank was still able to convince participants that they stole money after showing them a doctored video in which it appeared that they stole. *Id.* After watching the video, while knowing that they did not steal, participants would confess to taking money from the bank. *Id.*

towards admitting audiovisual evidence[63] because it is persuasive and easy for jurors to comprehend[64] and remember.[65] An example of this use of the deepfake defense was displayed in the defense of January 6 Insurrectionist Guy Reffitt.[66] As described elsewhere, Reffitt's lawyer attempted to defend him against the federal charges related to his attack on the United States Capital on January 6, 2022, by using the deepfake defense.[67] He argued to the jury that the video evidence, which had been authenticated and admitted in the trial, clearly showed Reffitt carrying a gun, storming the security barriers on the steps of the Capital, and leading the attack on the Capital police was a deepfake.[68] Notably, Reffitt's

[63] *See* Granot, Balcetis, Feigenson & Tyler, *supra* note 62, at 93 (describing the holding in United States v. Watson, 483 F.3d 828 (D.C. Cir. 2007)). Prosecutors striking blind persons from the jury emphasizes the idea that one must be able to see in order to fully comprehend all of the evidence in a case. *Id.* Visual evidence is so highly regarded in the justice system that it is seen as a way to mitigate juror bias towards other pieces of evidence in a case. *Id.*

[64] Studies have demonstrated that video evidence powerfully affects human memory and view of reality. Kimberley A. Wade, Sarah L. Green & Robert A. Nash, *Can Fabricated Evidence Induce False Eyewitness Testimony?*, 24 APPLIED COGNITIVE PSYCH. 899, 900 (2010). In 2010, researchers at the University of Warwick conducted a study illustrating the psychological effect that video has on reconstructing personal observations. *Id.* The researchers placed sixty college students in a room to engage in a computerized gambling task. *Id.* at 901–02. Following completion of the task, researchers individually showed each subject digitally altered video depicting a co-subject cheating, when in fact none of the subjects had actually cheated. *Id.* at 903–04. Nearly half of the subjects were willing to testify that they had personally witnessed a co-subject cheating after seeing the fake video; only one in ten was willing to testify to the same effect after the researcher merely told the subject about the cheating, rather than showing the fake video evidence. Hadley Leggett, *Fake Video Can Convince Witnesses to Give False Testimony*, WIRED (Sept. 14, 2009), https://www.wired.com/2009/09/falsetestimony [https://perma.cc/M88G-8TKJ]. "[R]esearchers emphasized that no one should testify unless they were 100 percent sure they had seen their partner cheat." *Id.*

[65] Studies demonstrate that jurors who hear oral testimony along with video testimony are 650% more likely to retain information. *See* Karen Martin Campbell, *Roll Tape—Admissibility of Videotape Evidence in the Courtroom*, 26 U. MEM. L. REV. 1445, 1447 (1996) (providing statistics on how jurors retain videotaped information at trial). Jurors who received visual testimony were 100% more likely to retain information than jurors who received only oral testimony. *Id.*; *see also* 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, 185 (2009) (citing statistics on the impact of visual evidence on jurors). "Jurors often are bored, confused, and frustrated when attorneys or witnesses try to explain technical or complex material" and having visual aids can help them retain information much better. *Id.* at 184–85. Jurors can retain up to 85% of information visually and in contrast only retain about 10% of what they hear. *Id.* at 185.

[66] *See* Verkouteren, *supra* note 6.

[67] *Id.*

[68] *Id.*

counsel presented no witnesses or other evidence during the trial—the deepfake defense was the only argument he made to the jury on Reffitt's behalf.[69]

The second context in which the deepfake defense arises occurs when it is asserted during the trial when counsel objects to the admission of authentic evidence, claiming it was altered. Such an objection poses a significant challenge to the effective prosecution of a case. If, for example, in a criminal case, audiovisual evidence is admitted, but the defense argues that it is inauthentic, the burden of proof for the prosecution has now become one of proving a negative rather than a positive—the opposite role to which many prosecutors are accustomed.[70] This not only puts prosecutors in an unusual position; it may prove an impossible burden given how difficult it is to prove that something is not a deepfake beyond a reasonable doubt.[71] This variant of the deepfake defense was displayed during the high-profile criminal trial of Kyle Rittenhouse, the teenager charged with shooting and killing several demonstrators during a racial justice protest in Kenosha, Wisconsin, in 2020.[72] During the trial, Rittenhouse's lawyer objected to the prosecution's request to play surveillance footage of the shooting for the jury on an Apple iPad.[73] The footage showed Ritten fatally shooting one of the victims.[74] Still, the defense objected that the built-in features on the iPad designed to zoom in and enlarge the images could not be trusted because Apple's artificial intelligence manipulated and altered the image.[75] Counsel proclaimed: "this is Apple's iPad programming creating what it thinks is there, not what necessarily is there."[76] Convinced by the argument, the trial judge did not permit the prosecutor to play the video without demonstrating that the use of the iPad did not alter the

---

[69] *Id.*

[70] *See* AGNES E. VENEMA & ZENO J. GERADTS, DIGITAL FORENSICS, DEEPFAKES, AND THE LEGAL PROCESS 17 (2020).

[71] *Id.* at 16–17.

[72] Kathleen Foody, *Tech Disputes at Rittenhouse Trial Not New Issue for Courts*, ASSOCIATED PRESS (Nov. 12, 2021), https://apnews.com/article/kyle-rittenhouse-technology-wisconsin-kenosha-homicide-b561bef68dc6aadaadc9b45a1bd93a19 [https://perma.cc/AW4L-PCQY].

[73] *Id.*

[74] *Id.*

[75] *Id.*

[76] Sean Hollister, *Judge Buys Rittenhouse Lawyer's Inane Argument that Apple's Pinch-to-Zoom Manipulates Footage*, VERGE (Nov. 10, 2021), https://www.theverge.com/2021/11/10/22775580/kyle-rittenhouse-trial-judge-apple-ai-pinch-to-zoom-footage-manipulation-claim [https://perma.cc/KTT4-FVQ9].

image.[77] The deepfake defense has also been deployed in cases involving child pornography[78] and assault with attempt to murder.[79]

C. *Lawyers' Use of the Deepfake Defense Is Wrong and Dangerous*

As the Reffitt and Rittenhouse cases show, counsel deploy the deepfake defense in trials to sway the jury and trial judges; thus, the threat presented by the deepfake defense is real. But is that threat dangerous to our legal system? And is counsel's use of the deepfake defense wrongful? Why is an attorney's use of the deepfake defense any different from other misrepresentations and lies that attorneys are permitted to assert on behalf of their clients?[80]

Deepfakes pose unique dangers to our legal system because they invite the viewer—in this context, the trier of fact, and more specifically, the jury or trial judge—to choose their truth based not merely on the evidence presented in the trial but also based on their individual perceptions of truth.[81] Thus, deepfakes, unlike other types of evidence that may be challenged as false or fabricated, are different. Jurors are bombarded by false images and messaging about fake news in a way they are not about other types of new or scientific evidence, such as DNA, X-rays, and photographs; deepfakes raise existential questions about reality that other types of evidence do not pose.[82]

The deepfake phenomenon has been incubated by and through the internet, which has elevated digital images as sources of information. In particular, social media is now the primary source of news and information for most Americans.[83] Social media elevates audiovisual content—and videos in particular—over other content formats.[84] But social media also spreads disinformation and

---

[77] Foody, *supra* note 72.

[78] *See* Schaffer v. Shinn, No. CV 20-08157 PCT JAT (CDB), slip op. at 7 (D. Ariz. 2021) (defendant attacked sufficiency of the evidence supporting sentencing enhancement arguing that the pornographic image was a deepfake).

[79] *See* People v. Smith, 969 N.W.2d 548, 565–66 (Mich. Ct. App. 2021) (defendant challenged the admission of Facebook posts belong to others which purportedly included his image and gang moniker, suggesting that they were fake).

[80] *See infra* Part III.B (describing the range of attorney conduct permitted under the rules of professional responsibility including acceptable lies and misrepresentations); *see also* Bruce A. Green & Rebecca Roiphe, *Lawyers and the Lies They Tell*, 69 WASH. U. J.L. & POL'Y 37, 70–75 (2022) (describing the various contexts in which lawyer lies are expected under the rules of professional conduct).

[81] Delfino, *Deepfakes on Trial*, *supra* note 13, at 345–46.

[82] *Id.* at 345.

[83] *See* 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].

[84] *See, e.g.*, 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].

deepfakes and drives news coverage of deepfakes.[85] And, as the public discovers the potential to be fooled by deepfakes, they will doubt authentic videos. The mistrust and instability that resulted illustrate the public's inability to gauge authenticity in the age of deepfakes[86] and the danger posed by that skepticism.[87]

As public knowledge of deepfakes continues to grow and people become increasingly skeptical about the credibility of audiovisual images, jurors will be primed to doubt the authenticity of even real audio and video content.[88] This juror skepticism may lead to plummeting confidence in video evidence absent a sponsoring witness, even if a judge authenticates it before reaching the jury.[89]

---

[85] *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.moth-erjones.com/politics/2019/03/deepfake-gabon-ali-bongo/ [https://perma.cc/9VSS-FD8G].

[86] Few empirical studies on the human ability to detect deepfakes currently exist, one behavioral experiment study published in late 2021 conducted by researchers from the Center for Humans and Machines, Max Planck Institute for Human Development, and University of Amsterdam School of Economics found that the study participants could not reliably detect deepfakes even after they were taught about them. Nils C. Köbis, Barbora Doležalová & Ivan Soraperra, *Fooled Twice: People Cannot Detect Deepfakes but Think They Can*, ISCIENCE, Oct. 29, 2021, at 1 (research study demonstrating people are not able to detect deepfakes reliably). The study demonstrates that people are biased toward mistaking deepfakes for authentic videos and overestimate their abilities to detect deepfakes. *Id.* at 11. Furthermore, researchers observed that participants adopted a "seeing-is-believing" heuristic for deepfake detection. *Id.* at 11. This combination renders people particularly susceptible to being influenced by deepfake content. *Id.* at 1. The study concludes that "detection of deepfakes is not a matter of lacking motivation but inability." *Id.* at. 9.

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

[88] *Id.*; *see* Matt Reynolds, *Courts and Lawyers Struggle with Growing Prevalence of Deepfakes*, AM. BAR ASS'N J. (June 9, 2020), https://www.abajournal.com/web/article/courts-and-lawyers-struggle-with-growing-prevalence-of-deepfakes [https://perma.cc/7ALF-PKVW]; Brown, *supra* note 54, at 25–26 (asserting that even if a realistic deepfake is publicly identified as fake it is unclear that the public would believe it).

[89] *See* Mika Westerlund, *The Emergence of Deepfake Technology: A Review*, 9 TECH. INNOVATION MGMT. REV. 39, 42–43 (2019) (describing how the public may begin to distrust authorities deemed reliable in the past because of deepfakes); Nicholas Mirra, *Putting Words in Your Mouth: The Evidentiary Impact of Emerging Voice Editing Software*, 25 RICH. J.L. & TECH. 1, 3 (2018) (cautioning that courts must be prepared for how "new technology may threaten existing and well established forms of evidence"); Holly Kathleen Hall, *Deepfake Videos: When Seeing Isn't Believing*, 27 CATH. U. J.L. & TECH. 51, 58 (2018) (contending that video may lose its value because "[t]he same accountability video that brings action can now be abused in a number of ways"); Drew Harwell, *Top AI Researchers Race to Detect 'Deepfake' Videos: 'We Are Outgunned'*, WASH. POST (July 12, 2019), https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/ [https://perma.cc/3N8Y-C484] (warning how the public may begin to generally distrust video footage because "[i]t's too much effort to figure out what's real and what's not").

Moreover, juror skepticism is problematic because it may allow bad actors to escape accountability simply because the jury has no means to prove that a piece of content is not, in fact, a deepfake.[90] Ultimately, if this tactic of challenging authenticity is used successfully in multiple cases, video evidence will lose its persuasive power, and taken far enough, it could degrade public trust in the very institution of courts themselves.[91]

Lawyers play an important role in society's confidence in legal institutions and in helping the jury interpret the evidence presented at trial.[92] The emergence of deepfakes presents a new opportunity to challenge the authenticity of audio-visual evidence that did not exist before 2018.[93] Exploiting the existence of deepfakes, in civil or criminal cases, like Guy Reffitt's, where counsel knows the video images are genuine, is a species of "false defense" that law should not countenance.[94]

Attorneys employ a "false defense" when they assert a false conclusion, draw a "false inference," or advance an alternate theory from truthful evidence (or, more accurately, evidence or testimony not known by defense counsel to be false).[95] While an attorney advancing an alternate theory need not directly make a false statement, the goal is to convince the fact-finder of a conclusion that the defense counsel knows to be false; in other words, "weakening the persuasiveness" of the prosecution's case by raising possibilities favorable to the client.[96] Professor David Luban characterizes using a "false inference" as "[s]uggesting reasonable doubts," which he points out often elides in practice into knowingly advancing a false conclusion, impeaching a witness known to be testifying truthfully, or "[s]aying things that are literally true but drastically misleading."[97] Guy Reffitt's lawyer's argument to the jury that the evidence against Reffitt was a deepfake[98] is an example of a "false inference" of Professor Luban's argument.

---

[90] Pfefferkorn, *supra* note 55, at 269–70.

[91] *Id.* at 270–71.

[92] *Id.* at 267–73.

[93] *Id.* at 270–71 (comparing deepfakes to the "CSI effect" whereby jurors demanded sophisticated scientific evidence because they knew it existed due to the hit TV show).

[94] As used in this Article, the deepfake defense constitutes lawyers lying, that is making statements the lawyer knows to be false intending that others believe these statements. This definition does not include instances where a lawyer makes a false statement in the honest but mistaken belief the statement is true or with doubts about its veracity. It also excludes lawyers making literally true but misleading statements, making misleading omissions, deliberately failing to correct others and mistaken beliefs.

[95] John B. Mitchell, *Reasonable Doubts Are Where You Find Them: A Response to Professor Subin's Position on the Criminal Lawyer's "Different Mission*," 1 GEO. J. LEGAL ETHICS 339, 340 (1987).

[96] *Id.* at 346 (emphasis omitted).

[97] *See* David Luban, *Are Criminal Defenders Different?*, 91 MICH. L. REV. 1729, 1759-60 (1993).

[98] Verkouteren, *supra* note 6.

Two categories of false inference exist: (1) those intended to show how the prosecution has failed to carry its burden of proof; and (2) inferences "used for their probative value."[99] The difference between these two forms of false inference is slight but meaningful. An example of the former is a defense attorney who argues to the jury that, based on the prosecution's evidence, it cannot be known beyond a reasonable doubt whether the (guilty) defendant committed the crime.[100] This form of false inference is widely accepted as an appropriate way to put the government to its proof because the defense counsel simply is illustrating with particularity how the prosecution has failed to make its case.[101] The latter form of false inference—one relied upon by defense counsel for its purported truth—by its terms, seeks to deceptively convince the fact-finder of its truthfulness by claiming the evidence that the defense counsel knows to be truthful is false.[102] The deepfake defense falls into this latter category of false inference, which should not be permitted because it undermines the jury's ability to perceive and understand the evidence and to collectively determine the truth.

As the harms caused by deepfakes have exploded into the public square, policymakers have shifted into containment and response mode—attempting to prevent, mitigate, and punish the abuse of deepfake technology for harmful purposes.[103] However, thus far, few federal and state laws have targeted

---

[99] Ellen Yankiver Suni, *Who Stole the Cookie from the Cookie Jar?: The Law and Ethics of Shifting Blame in Criminal Cases*, 68 FORDHAM L. REV. 1643, 1652, 1656 (2000).

[100] *See id.* The goal is to convince the factfinder that a false narrative favorable to the defendant cannot be disproven beyond a reasonable doubt by affirmatively pointing out specific ways in which the government has failed to prove its case. Mitchell, *supra* note 95, at 340.

[101] Suni, *supra* note 99, at 1656.

[102] *Id.* at 1658–59. As an example, consider an attorney who exploits a deficiency in the prosecution's evidence by arguing that a truthful eyewitness did not actually observe the defendant committing the crime. Indeed, in practice this tactic often occurs when a defense attorney subjects a truthful witness to cross-examination intended to impugn the witness's credibility or challenges evidence that the attorney knows to be truthful.

[103] *See, e.g.*, Andrew J. Grotto, *Written Testimony of Andrew J. Grotto on "Cybersecurity and California Elections*,*"* STAN. CTR. FOR INT'L SEC. & COOP. (Mar. 7, 2018) (testimony of Andrew Grotto before California Legislature on elections, redistricting, and constitutional amendments), https://cisac.fsi.stanford.edu/docs/andrew-grotto-testimony-cybersecurity-and-california-elections [https://perma.cc/5AM8-NKYC]; Kaveh Waddell, *Lawmakers Plunge into 'Deepfake' War*, AXIOS (Jan. 31, 2019), https://www.axios.com/deepfake-laws-fb5de200-1bfe-4aaf-9c93-19c0ba16d744.html [https://perma.cc/QA5U-ZBZ6] (reporting that federal legislators had begun "invit[ing] legal scholars to privately brief their staff on deepfakes"). Scholars likewise have explored existing civil remedies and criminal laws that could be applied to redress those harms and proposed new laws and legal frameworks to constrain bad actors. Chesney & Citron, *supra* note 19, at 1777; Delfino, *Pornographic Deepfakes*, *supra* note 12, at 903–37.

deepfakes.[104] And because of the law's limited ability to contain deepfakes, the increasing ease with which deepfakes are created, and the frequency of their appearance on the internet, deepfakes and the use of the deepfake defense are not likely to go away any time soon.

The next part of this Article examines the existing legal means under the rules of procedure, substantive law, and ethics to address the challenges of the deepfake defense in legal proceedings and argues that current means are insufficient to address the impact in the courtroom.

III. CURRENT MECHANISMS TO ADDRESS THE CHALLENGES OF THE DEEPFAKE DEFENSE IN LEGAL PROCEEDINGS AND WHY THEY ARE INADEQUATE

The law offers various methods to deal with lawyers' conduct in court proceedings, and more to the point, lawyers' objections to and arguments about whether evidence is genuine or fake. Some of these mechanisms might be applied to the deepfakes defense. Specifically explored in this part, courts, bar regulators, and lawyers will look to the rules of procedure, ethics, and substantive law to address the challenges presented by the deepfake defense. However, before addressing the current methods that may be applied and their respective inadequacies, this part begins by locating the deepfake defense within the historical context of the lawyers' efforts to manipulate and corrupt legal proceedings by offering false arguments. This context is important because it informs the understanding of how the current law would address the deepfake defense and underscores the argument that the contemporary legal frameworks are insufficient to address the harms of the deepfake defense.

A. *Historical Context—Legal Responses to Lying Lawyers and Its Lessons for the Deepfake Defense*

When the ancient Greek Athenian[105] *logographoi*—the forerunners of modern trial lawyers—prepared oral arguments for litigants to deliver[106] to the Greek juries, they used guile and falsehood in their arguments.[107] The Romans

---

[104] The landscape of existing laws addressing deepfakes is beyond the scope of this article, but is explored elsewhere. *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 298–313.

[105] The ancient Greeks had a reputation as a "nation of lawyers," which inspired Aristophanes' play Wasps. Richard Underwood, *False Witness: A Lawyer's History of the Law of Perjury*, 10 ARIZ. J. INT'L & COMP. L., 215, 230 n.65 (1993); 1 ARISTOPHANES, THE WASPS (Benjamin Rogers trans., Loeb Classical Library 1982) (c. 422 B.C.E.).

[106] *Logographoi* were litigants speech-writers; they were not permitted to "appear" in legal proceedings on behalf of the litigants. Underwood, *supra* note 105, at 230.

[107] *Id.* The speeches were not confined to the evidence, in any modern sense, and the diecasts (jurors) "were at liberty to base their verdict on the unconfirmed statements of a

followed suit.[108] Though the Romans had a professional class of lawyers who should have been interested in fairness in litigation and candor to the courts and thus should have urged the adoption of legal rules or ethical practices, they did not.[109] Instead, Roman orators played shamelessly on the sympathies of the factfinders.[110] The secular law offered no response to this conduct—no code of ethics or legal rules existed to reign in these behaviors.[111] Instead, in the ancient western world, lying and fabrication in legal proceedings was "a matter which from first to last [was] left to the vengeance of the gods, and the law never threaten[ed] secular penalties against the offender."[112]

The regulation of lawyers that currently exists reflects a tension—not the tension of ancient times between earthly regulators and the wrath of God or the punishment of ecclesiastical authorities, but instead over who should regulate lawyers' conduct in legal proceedings and how much regulation is enough. By the end of the Middle Ages, the regulation of deceptive practices by lawyers came into the fold of the courts.[113] The modern legal profession arose in the twelfth and thirteenth centuries in England.[114] In 1292, to protect the public, the King of England placed the regulation of lawyers' conduct in the hands of the courts.[115] Chapter 29 of the First Statute of Westminster of 1275 made "deceit or collusion in the king's court, or consent unto it, in deceit of the court" on the part of a lawyer punishable by imprisonment for a year and a day.[116] In addition, the guilty lawyer also lost the right "to plead in that court for any man."[117] The

---

speaker if they deemed them trustworthy, or upon their own independent knowledge of the case." *Id.*

[108] *See id.* at 231.

[109] *See id.* Cicero lamented that the leading Roman advocates won cases through the use of politics, influence peddling, and the bribery of judges, rather than as a result of their legal ability. CICERO, AGAINST VERRES 105, 107 (T. E. Page, E. Capps & W.H.D. Rouse eds., Loeb Classical Library 1978) (c. 70 B.C.E.).

[110] Underwood, *supra* note 105, at 233. Roman advocates were "not expected to do even lip-service to the testimony before the court." 2 JAMES LEIGH STRACHAN-DAVIDSON, PROBLEMS OF THE ROMAN CRIMINAL LAW 121 (1912).

[111] Underwood, *supra* note 105, at 232.

[112] 1 JAMES LEIGH STRACHAN-DAVIDSON, PROBLEMS OF THE ROMAN CRIMINAL LAW 41, 48 (1912). Cicero described the punishment for such conduct as destruction from Heaven, "but from man only blame and contempt" would follow. *Id.* at 49 (citing CICERO, DE LEGIBUS, II.9.22). Until the middle ages, even perjury was not considered a criminal offense. Underwood, *supra* note 105, at 229, 236. Secular law deemed it a spiritual offense, placing the obligation to prevent and punish this conduct on the church. *Id.* at 240.

[113] John S. Bradway, *Moral Turpitude as the Criterion of Offenses that Justify Disbarment*, 24 CALIF. L. REV., 9, 9 (1935).

[114] *Id.*

[115] *Id.*

[116] The Statutes of Westminster 1275, 3 Edw. 1, c. 29 (Eng.).

[117] *Id.* at 12–13. According to Professor Jonathan Rose, the "chapters were directed at abuses that impacted the operation of the judicial system" including "champerty, extortion, bribery, abuse of official power, maintenance, and abusive litigation practices by royal and court officials, lawyers, and individual litigants." Jonathan Rose, *The Legal Profession in*

phrase "deceit or collusion" was given an expansive interpretation.[118] And through the regulation of lawyers by the judiciary and the bar in England, this time has been characterized as a "stable adjustment."[119] This stability in the bar did not cross the Atlantic Ocean to American lawyers who practiced during the colonial era. But there were specific attempts by colonial legislators to address attorney deceit—at least some of the American colonies regulated admission to the practice of law by statute, which often contained oaths of office.[120] These oaths sometimes addressed attorney deceit. For example, a 1708 Connecticut statute contained an attorney oath prohibiting an attorney from doing any "falsehood" in court and bringing a "false or unlawful suit."[121]

Later the regulation of lawyers was shared among the courts, bar associations, and state legislatures. This model of shared governance persists today.[122] In addition, to the shared governance of lawyers, the degree to which lawyers' conduct should be regulated and, relevant here, the rules and laws that dictate when and how a lawyer may use fabrication, deception, or urge the deepfake defense, reflects the complex role that lawyers have been assigned in society and the duties and obligations imposed on them. Tension exists among regulators, the courts, and within the rules and laws that govern lawyers' behavior because of competing obligations borne from the adversarial and

---

Medieval England: A History of Regulation, 48 SYRACUSE L. REV. 1, 53 (1998). Rose's review of the commentary occurring contemporaneously and subsequent to the enactment of the statute led him to conclude that the primary concern of chapter 29 was in addressing lawyer misconduct "because of its negative impact on the justice system." *Id.* at 56.

[118] Rose, *supra* note 117, at 58. According to Rose:

> [T]he cases involved a wide range of lawyer misconduct. The covered conduct included forgery of writs; altering, damaging or removing official documents; conflict of interest and other breaches of client loyalty; false statements to the court, the client, the opponent, and in pleadings and other documents; acting as an attorney without proper authority, or continuing to act after removal; failing to act or premature termination of representation; antagonizing judges by unconvincing arguments, overzealousness, or not speaking in good faith; defective pleadings and documents; unjustified initiation or continuation of litigation, and repleading issues; and misconduct in the lawyer's own litigation or business dealings.

Id. at 61.

[119] Bradway, *supra* note 113, at 11.

[120] *See* Carol Rice Andrews, *Standards of Conduct for Lawyers: An 800-Year Evolution*, 57 SMU L. REV. 1385, 1417 (2004).

[121] JOSIAH HENRY BENTON, THE LAWYER'S OFFICIAL OATH AND OFFICE 42 (1909); *see also id*. at 70–72 (1791 New Hampshire statute containing oath of office with similar language).

[122] *See id.* at 12 (explaining that the question of state versus professional control over the bar and regulating lawyers' action still exists).

judicial system's dual aims of seeking truth and protecting individual rights.[123] On the one hand, lawyers must apply their skills and effort to zealously advocate and promote their client's causes and protect their confidence.[124] As advocates and fiduciaries, lawyers use rhetoric to spin the truth on behalf of clients.[125] And in some instances, as discussed in the following sections, the rules of law and professional ethics give lawyers significant leeway to engage in that deception.[126]

On the other hand, lawyers, as officers of the court, also have duties to others within the justice system, including duties of candor to the courts and fairness to the opposing counsel, parties, witnesses, and others.[127] They are also responsible for not engaging in conduct involving dishonesty, fraud, deceit, or misrepresentation.[128] These competing obligations are baked in the rules of procedure[129] and substantive law.[130] They are most apparent in the ethical rules.[131] According to the Preamble to the ABA Model Rules governing professional conduct, lawyers must balance their duties as "a representative of clients, an officer of the legal system, and a public citizen having special responsibility for the quality of justice."[132] Thus, lawyers play a "vital role in the preservation of society" and must cultivate "democratic competence."[133]

The tension between the competing duties of lawyers coupled with the shared regulation of the legal profession by the courts and state bar associations has added a layer of complexity to the law that governs lawyers' conduct. As our society transforms into one dominated by new complicated technologies like AI and deepfakes, courts and bar regulators must consider whether the existing rules of procedural and ethics and substantive law are sufficient to address new problems like the deepfake defense. The rules and laws these regulators apply are examined in the next part of this Article. And as described in the next section, although we no longer live in the days of Greece or Rome when lawyers had free reign to lie, mislead, and deceive, there is a laxity built into the existing law that appears to allow lawyers to use the deepfake defense with impunity.[134]

---

[123] Renee Knake Jefferson, *Lawyer Lies and Political Speech*, 131 YALE L.J. F., 114, 116, 133 (2021).

[124] *See* Green & Roiphe, *supra* note 80, at 38–39.

[125] *Id.* at 39.

[126] Jefferson, *supra* note 123, at 116, 125–27.

[127] *See* discussion *infra* Part III.B.2. *See generally* MODEL RULES OF PRO. CONDUCT r. 3.1, 3.3, 4.1 (AM. BAR ASS'N 2021).

[128] *See* discussion *infra* Part III.B.2. *See generally* MODEL RULES OF PRO. CONDUCT r. 8.4 (AM. BAR ASS'N 2021).

[129] *See* discussion *infra* Part III.B.2; FED R. CIV. P. 11.

[130] *See* discussion *infra* Part III.B.2. *See generally* MODEL RULES OF PRO. CONDUCT r. 3.1, 3.3, 4.1, 8.4 (AM. BAR ASS'N 2021).

[131] *See* discussion *infra* Part III.B.2.

[132] MODEL RULES OF PRO. CONDUCT Preamble ¶ 1 (AM. BAR ASS'N 1983).

[133] Jefferson, *supra* note 123, at 133.

[134] *See* discussion *infra* Part III.B.1.

B. *Rules of Procedure, Substantive Law, and Ethical Rules and Their Limitations*

This section identifies the available legal and ethical responses to address the attorney's use of the deepfake defense. It also explains that despite the variety of options at the disposal of courts and disciplinary authorities to address the deepfake defense, limitations exist, and current regulatory approaches to the deepfake defense are inadequate.

1. *Current Legal Approaches to Address the Deepfake Defense and Why They Fall Short*

Courts have several options to address attorney misconduct and abuses of the trial processes. The most basic is a court's inherent power to sanction.[135]

a. *Courts' Inherent Power to Sanction Misconduct and Its Limitations*

Courts have broad inherent power to control litigation conduct.[136] For example, in the federal context, the Supreme Court has held that district courts have the inherent power to sanction "to achieve the orderly and expeditious disposition of cases."[137] However, courts differ regarding the kinds of sanctions they can impose. For example, federal district courts are permitted to impose procedural (e.g., exclusion of evidence or dismissal) and monetary sanctions for misconduct.[138] Included in the latter are compensatory penalties and attorney's fees.[139]

The Supreme Court has ruled that "a federal court [is not] forbidden to sanction bad faith conduct by means of the inherent power simply because that conduct could also be sanctioned under the statute or the Rules."[140] This is

---

[135] Daniel J. Meador, *Inherent Judicial Authority in the Conduct of Civil Litigation*, 73 TEX. L. REV. 1805, 1805–06 (1995).

[136] *See id.* at 1805–07 (describing the history and development of the law relating to the court's inherent authority to control litigation conduct); Jaymie L. Roybal, Note, *Permission to Punish: Sanctions Without Boundaries*, 46 N.M. L. REV. 217, 220–22 (2016); Chambers v. NASCO, Inc., 501 U.S. 32, 49–50 (1991).

[137] Link v. Wabash R.R. Co., 370 U.S. 626, 630–31 (1962); Big Yank Corp. v. Liberty Mut. Fire Ins. Co., 125 F.3d 308, 313 (6th Cir. 1997) ("A district [court has the] inherent authority to award fees when a party litigates in bad faith, vexatiously, wantonly, or for oppressive reasons." (internal quotation marks omitted)).

[138] *See* Meador, *supra* note 135, at 1806.

[139] *See* Goodyear Tire & Rubber Co. v. Haeger, 137 S. Ct. 1178, 1186 (2017); Roadway Express, Inc. v. Piper, 447 U.S. 752, 766–67 (1980). However, federal courts' inherent sanction power does not permit the levying of punitive damages. *See Goodyear*, 137 S. Ct. at 1186.

[140] *Chambers*, 501 U.S. at 50. In contrast, the inherent authority of California superior courts to impose monetary sanctions is more limited than that of the federal courts. *See* Bauguess v. Paine, 586 P.2d 942, 948–49 (Cal. 1978); Bryan v. Bank of Am., 86 Cal. App.

important for two reasons. First, federal courts may invoke their inherent power to sanction without grounding the decision in a rule or statute. Second, the federal court's inherent power to sanction will sometimes overlap with the sanctioning power granted to them by statute or rule. A recent example of the court's reliance upon its inherent authority was displayed in *King v. Whitmer*.[141] There the plaintiffs argued that the 2020 election of Joseph Biden to United States President resulted from a conspiracy and voter fraud that altered the Michigan voting results and that the Michigan results should therefore be decertified.[142] The *King* court found these claims to be "fantastical."[143] Citing its "inherent authority," the court imposed sanctions on the attorneys in that case for promoting falsehoods, noting that they lacked any basis in fact or evidence.[144] Sanctions serve many goals in addressing attorney deceit, including deterrence, punishment, and promoting respect for the legal process.[145] In addition, judges may tailor sanctions on an individual basis as appropriate.[146]

Although the federal court's inherent sanctioning power is broad, not limited by an overlapping statute or rule, some circuit courts have encouraged district courts to seek sanctioning authority in a statute or rule before grounding it in their inherent power.[147] The Seventh Circuit is notable for holding that a district court's inherent sanctioning authority "is a residual authority, to be exercised sparingly, to punish misconduct (1) occurring in the litigation itself . . . and (2) not adequately dealt with by other rules . . . ."[148] This is as descriptive of how district courts use their inherent sanction power as it is prescriptive of how they should use it; district courts often only invoke their inherent sanctioning authority when the misconduct in question cannot be sanctioned under statute or rule.[149]

4th 185, 187 (2001). Typically, if a California superior court wishes to impose monetary sanctions, it must find the power to do so in a statute or court rule. *Bauguess*, 586 P.2d at 948–49.

[141] King v. Whitmer, 556 F. Supp. 3d 680, 689–90 (E.D. Mich. 2021).

[142] *Id.* at 690.

[143] *Id.* at 732 n.82.

[144] *Id.* at 689–90.

[145] Thomas C. Tew, *Electronic Discovery Misconduct in Litigation: Letting the Punishment Fit the Crime*, 61 U. MIA. L. REV. 289, 306, 322–23 (2007) (arguing that "there are many opportunities for improvement in the Rules [of Civil Procedure]" to address discovery abuses).

[146] *See, e.g.*, Reilly v. NatWest Mkts. Grp., 181 F.3d 253, 267 (2d Cir. 1999) (noting ability of district judges to tailor sanctions against those who spoliate evidence); Tew, *supra* note 145, at 323 (noting courts have broad discretion in crafting sanctions for discovery abuse).

[147] *See, e.g.*, *In re* Prudential Ins. Co. Am. Sales Prac. Litig. Agent Actions, 278 F.3d 175, 189 (3d Cir. 2002); Zapata Hermanos Sucesores, S.A. v. Hearthside Baking Co., 313 F.3d 385, 390–91 (7th Cir. 2002).

[148] *Zapata Hermanos Sucesores, S.A.*, 313 F.3d at 390–91.

[149] *See, e.g.*, Perry v. S.Z. Rest. Corp., 45 F. Supp. 2d 272, 276 (S.D.N.Y. 1999); Thomason v. Norman E. Lehrer, P.C., 182 F.R.D. 121, 132 (D.N.J. 1998); Weeks

Although judicial sanctions are available, they may not be an effective deterrent because courts have been reluctant to use their inherent authority to sanction counsel for their transgressions in legal proceedings.[150] This tepid approach to the exercise of its inherent power to reign in attorneys who use the deepfake defense likely means that unless the deepfake defense is prohibited by a procedural rule, statute, or the rules of ethics, the courts may not engage their inherent power to sanction attorneys who use it.

Some doubt the effectiveness of judicial sanctions in dealing with abusive litigation and discovery-related misconduct.[151] Some commentators have criticized courts for their unwillingness to impose more severe sanctions.[152] In addition, because one of the goals of judicial sanctions is to restore the innocent party to the position they would have been in, but for the other's side misconduct,[153] the preferred remedy in many cases often involves nonmonetary sanctions such as adverse inferences regarding evidence or pleadings. Thus, the prejudiced party often goes without compensation for the added time, expense, and anxiety the other side's misconduct caused.[154] Critics have also questioned whether there is anything genuinely punitive about even the most severe nonmonetary sanctions. For example, Professor Charles R. Nesson has asserted that parties only destroy or conceal evidence because they believe it will damage their cases.[155] Therefore, assuming the spoliator would have lost anyway,

---

Stevedoring Co., v. Raymond Int'l. Builders, Inc., 174 F.R.D. 301, 306 (S.D.N.Y. 1997); Mahoney v. Yamaha Motor Corp. U.S.A., 290 F.R.D. 363, 368 (E.D.N.Y. 2013); Sommerfield v. City of Chicago, 252 F.R.D. 407, 419 n.8 (N.D. Ill. 2008).

150 *See* sources cited *supra* note 149.

151 *Compare* Bell Atl. Corp. v. Twombly, 550 U.S. 544, 559 (2007) ("[T]he common lament that the success of judicial supervision in checking discovery abuse has been on the modest side."), *and* Robert W. Gordon, *The Ethical World of Large-Firm Litigators: Preliminary Observations*, 67 FORDHAM L. REV. 709, 736 (1998) ("Though perceptions differ, there seems to be some consensus that adversary excess is frequent, often not by any standard justifiable as zealous representation, and that many lawyers will indeed cross ethical lines when they think they can get away with it, which, because of the weakness of monitoring agents, they usually do."), *with* Peter A. Joy, *The Relationship Between Civil Rule 11 and Lawyer Discipline: An Empirical Analysis Suggesting Institutional Choices in the Regulation of Lawyers*, 37 LOY. LA. L. REV. 765, 811 (2004) (arguing that Rule 11 sanctions have generally been effective in deterring litigation misconduct).

152 *See* Charles R. Nesson, *Incentives to Spoliate Evidence in Civil Litigation: The Need for Vigorous Judicial Action*, 13 CARDOZO L. REV. 793, 793 (1991).

153 *See* Tew, *supra* note 145, at 323.

154 *See* Scott A. Moss, *Reluctant Judicial Factfinding: When Minimalism and Judicial Modesty Go Too Far*, 32 SEATTLE U. L. REV. 549, 560 (2009) (stating that "the trend in recent years [is] more judicial willingness to award attorneys' fees and other monetary sanctions on discovery motions," but noting that "fee awards or other monetary sanctions are nowhere near as 'common"' as it might seem); Virginia L.H. Nesbitt, *A Thoughtless Act of a Single Day: Should Tennessee Recognize Spoliation of Evidence as an Independent Tort?*, 37 U. MEM. L. REV. 555, 575–76 (2007) (stating, in context of destruction of evidence, that sanctions typically invoked do not adequately serve goal of compensation).

155 Nesson, *supra* note 152, at 801.

entering a default judgment leaves the spoliator where it would have been had the evidence been preserved.[156] Furthermore, it denies the jury the opportunity to see how damaging the evidence truly was, thus possibly preventing the imposition of punitive damages.[157]

Consequently, searching for a legal mechanism to curb the use of the deepfake defense requires an examination of other means to address attorney misconduct.

b. *Legal Procedural Mechanisms and Their Limitations*

In addition to the court's inherent power to sanction a party for bad behavior, a court may also employ various procedural rules and statutes to address attorney misconduct.

i. *FRCP Rule 11*

FRCP Rule 11, governing representations made to the court,[158] has undergone several substantive revisions since its original adoption with the rest of the Federal Rules of Civil Procedure in 1938.[159] It governs representations made in *signed* filings, including pleadings and motions and subsequent reiterations of representations made in signed filings.[160] Consequently, Rule 11 does not apply to oral representations made to the court that lack a genesis in a signed writing.[161] Thus, if a court seeks to sanction a representation made outside of any signed filing, it must look elsewhere to do so.

The current version of Rule 11 has several notable features. First, Rule 11 imposes a duty on parties to make an objectively reasonable inquiry into a factual claim before representing it in a signed filing.[162] Parties cannot make a representation in a signed pleading just because they feel like or wish it is true.[163] Some recent examples of violations of this requirement concern the attempt to overturn the 2020 Presidential election. For example, in *King v. Whitmer*, in addition to relying upon its inherent authority to punish the lawyer's misconduct, the court also found that because even a rudimentary investigation

---

[156] *Id.*

[157] *Id.* at 801–02.

[158] *See* FED. R. CIV. P. 11(b) (applies to any "pleading, written motion, or other paper—whether by signing, filing, submitting, or later advocating it").

[159] For a detailed history of the revisions to Rule 11, see generally William Schwarzer, *Rule 11 Revisited*, 101 HARV. L. REV. 1013 (1988); and 5A CHARLES ALAN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 1331 (4th ed. 2023).

[160] Although the original version of Rule 11 applied to discovery proceedings, the present version does not. *See* FED. R. CIV. P. 11(d) ("This rule does not apply to disclosures and discovery requests, responses, objections, and motions under Rules 26 through 37.").

[161] *Id.*

[162] *See* Bus. Guides, Inc. v. Chromatic Commc'ns Enters., 498 U.S. 533, 541 (1991).

[163] *See id.*

refuted the claims, the *King* court sanctioned the plaintiffs' attorneys under Rule 11.[164] Similarly, in *O'Rourke v. Dominion Voting Systems Inc.*, plaintiffs alleged defendants were members of a nationwide scheme to commit election fraud.[165] The *O'Rourke* court also found these claims to be baseless and easily refuted, thus sanctioning plaintiffs' attorneys for violating Rule 11.[166]

Second, Rule 11 prohibits lawyers from using a signed writing for "[an] improper purpose, such as to harass, cause unnecessary delay, or needlessly increase the cost of litigation."[167]

Third, Rule 11 requires that affirmative representations made in signed filings have factual support or at least will likely have such support.[168] Having support does not mean the affirmative representations must be true. Rather, it simply means that they must not be baseless.[169] The election conspiracy representations in *King* and *O'Rourke* were sanctionable not only because they failed the reasonable inquiry requirement but also because they were baseless, in that plaintiffs' attorneys had no evidence of their truth.[170]

Fourth, Rule 11 makes clear that the purpose of any sanctions is to deter future occurrences of the same misconduct, not to compensate the other parties.[171] However, courts have been equally clear that such deterrence must be balanced against the imperative placed on attorneys to represent their clients zealously.[172]

---

[164] *See* King v. Whitmer, 552 F. Supp. 3d 680, 703 (E.D. Mich. 2021). Plaintiff's attorneys in *King* argued that they satisfied the reasonable inquiry requirement because they spoke to other lawyers about the conspiracy allegations. The *King* court disagreed, holding that talking to other lawyers does not, alone, satisfy Rule 11's reasonable inquiry requirement. *See id.*

[165] *See* O'Rourke v. Dominion Voting Sys's., Inc., 552 F. Supp. 3d 1168, 1174 (D. Colo. 2021).

[166] *Id.* at 1208.

[167] FED. R. CIV. P. 11(b)(1); *see, e.g.*, Lipin v. Nat'l Union Fire Ins. Co. of Pittsburgh, Pa., 202 F. Supp. 2d 126, 140–41 (S.D.N.Y. 2002) (using litigation to delay proceedings in another case); Sec. & Exch. Comm'n v. Knoxville, LLC, 895 F. Supp. 192, 195 (E.D. Tenn. 1995) (sanctioning the delay in compliance with a subpoena); Elster v. Alexander, 122 F.R.D. 593, 604 (N.D. Ga. 1988) (sanctioning coercing a settlement); Whitehead v. Food Max of Miss., Inc., 332 F.3d 796, 796 (5th Cir. 2003) (sanctioning tactics intended to embarrass opposing party or counsel).

[168] FED. R. CIV. P. 11(b)(3) (requiring "factual contentions have evidentiary support or, if specifically so identified, will likely have evidentiary support after a reasonable opportunity for further investigation or discovery").

[169] *See id.*

[170] *See Whitmer*, 552 F. Supp. 3d at 703; *O'Rourke*, 552 F. Supp. 3d at 1174.

[171] FED. R. CIV. P. 11(c)(4) ("A sanction imposed under this rule must be limited to what suffices to deter repetition of the conduct or comparable conduct by others similarly situated.").

[172] *See, e.g.*, Operating Eng'rs Pension Tr. v. A-C Co., 859 F.2d 1336, 1344 (9th Cir. 1988) ("Rule 11 must not be construed so as to conflict with the primary duty of an attorney to represent his or her client zealously."); *see also* FED. R. CIV. P. 11 advisory committee's

Finally, where Rule 11 sanctions are found to be warranted, a court has discretion in deciding whether to impose them.[173] Thus, Rule 11 does not make sanctions mandatory.

ii. *FRCP Rule 59*

FRCP Rule 59 enables a district court to order a new trial after the factfinder has rendered a verdict.[174] Unlike Rule 11, such orders are not limited to misconduct involving signed filings. Thus, a court can order a new trial for misconduct that cannot be sanctioned under Rule 11.[175] However, Rule 59 makes explicit that any new trial order must be grounded in precedent.[176] A new trial after a jury trial may be ordered only "for any reason for which a new trial has heretofore been granted in an action at law in federal court."[177] Thus, one must look to case law to determine whether a case may be retried under Rule 59.

iii. *FRCP Rule 60*

FRCP Rule 60 provides a different remedy for a party aggrieved by another party's misconduct. While Rule 59 permits ordering a new trial, Rule 60 permits a court to vacate an adverse judgment.[178] Like Rule 59, the reasons a court may invoke Rule 60 are not limited to issues with signed filing. Among the reasons a court may order relief from judgment are "mistake, inadvertence, surprise, or excusable neglect"[179] and "fraud . . . , misrepresentation, or misconduct by an

---

note to 1983 amendment ("The rule is not intended to chill an attorney's enthusiasm or creativity in pursuing factual or legal theories.").

[173] FED. R. CIV. P. 11(c)(1) ("If, after notice and a reasonable opportunity to respond, the court determines that Rule 11(b) has been violated, the court *may* impose an appropriate sanction on any attorney, law firm, or party that violated the rule or is responsible for the violation." (emphasis added)).

[174] FED. R. CIV. P. 59(a)(1).

[175] *Id.* (a new trial can be ordered for "any reason" for which a new trial/rehearing has already been granted in federal court).

[176] *Id.*

[177] FED. R. CIV. P. 59(a)(1)(A) ("The court may, on motion, grant a new trial on all or some of the issues—and to any party . . . after a jury trial, for any reason for which a new trial has heretofore been granted in an action at law in federal court."). And a new trial after a bench trial may be ordered only "for any reason for which a rehearing has heretofore been granted in a suit in equity in federal court." Fed. R. Civ. P. 59(a)(1)(B) ("The court may, on motion, grant a new trial on all or some of the issues—and to any party . . . after a nonjury trial, for any reason for which a rehearing has heretofore been granted in a suit in equity in federal court.").

[178] FED. R. CIV. P. 60(b) ("On motion and just terms, the court may relieve a party or its legal representative from a final judgment, order, or proceeding for the following reasons . . . .").

[179] FED. R. CIV. P. 60(b)(1).

opposing party."[180] The concepts of surprise and fraud—where a lawyer knowingly makes a false representation to the court—is particularly relevant here. Suppose a party waits until trial to raise a defense, such as a deepfake defense. This surprise defense at trial may warrant the court granting the surprised party relief from an adverse judgment, especially since the surprised party will not have adequate time to prepare a rebuttal.[181] Suppose instead of surprising the opposing party with a previously unmentioned defense, a party's defense was dismissed by the court before trial, but the party ignores the dismissal and raises the defense during trial and argues to the jury that the evidence presented is a deepfake. Not only might such misconduct be grounds for FRCP Rule 11 sanctions, but it may also be grounds for Rule 60 relief from an adverse judgment for the opposing party.[182] If the party pressing the previously dismissed defense *knows* that the defense is baseless—that is, if the party knows that its factual bases are false—then such conduct may rise to the level of misrepresentation, providing additional grounds for a court to order sanctions and relief from judgment.[183]

iv. *Procedures in Criminal Cases*

The federal rules of criminal procedure do not contain an analog to the FRCP Rules 11, 59, or 60. Aside from the requirement that every person filing a motion or paper in court sign it,[184] there is nothing in the federal rules of criminal procedure providing a sanction for lawyers for acting in bad faith, nor do the rules address the lawyer's conduct at issue here, relying on a false defense.[185] Consequently, in a criminal trial, when the defendant baselessly but

---

[180] FED. R. CIV. P. 60(b)(3).

[181] *See, e.g.*, Knowles v. Mut. Life Ins. Co. of N.Y., 788 F.2d 1038, 1040 (4th Cir. 1986).

[182] *See* Schreiber Foods, Inc. v. Beatrice Cheese, Inc., 402 F.3d 1198, 1206 (Fed. Cir. 2005) ("The Seventh Circuit has held that a new trial on all issues may be granted as a form of sanction for attorney misconduct.").

[183] *See* Hesling v. CSX Transp., Inc., 396 F.3d 632, 641 (5th Cir. 2005) ("A party making a Rule 60(b)(3) motion must establish (1) that the adverse party engaged in fraud or other misconduct, and (2) that this misconduct prevented the moving party from fully and fairly presenting his case.").

[184] FED. R. CRIM. P. 49(b)(4) ("Every written motion and other paper must be signed by at least one attorney of record in the attorney's name—or by a person filing a paper if the person is not represented by an attorney. The paper must state the signer's address, e-mail address, and telephone number. Unless a rule or statute specifically states otherwise, a pleading need not be verified or accompanied by an affidavit. The court must strike an unsigned paper unless the omission is promptly corrected after being called to the attorney's or person's attention.").

[185] *See* Joshua A. Liebman, Note, *Dishonest Ethical Advocacy?: False Defenses in Criminal Court*, 85 FORDHAM L. REV. 1319, 1330–35 (2016) (discussing the range of conduct which qualify as false defenses in criminal cases, including putting the government to its proof, inviting the fact finder to draw false inferences from truthful evidence, arguing actual innocence, shifting the blame to others, perjury and knowingly submitting false

convincingly argues that the audio-video evidence against them is fabricated, the prosecution's options to respond are limited. The government can object to the arguments and try to have them stricken.[186] But the damage may already have occurred because it is unlikely the jury will be able to ignore the argument in their deliberations. In that case, the government may seek a mistrial.[187]

A mistrial can be a powerful deterrent to attorneys from presenting spurious defenses in criminal proceedings. If the motion is granted, the attorneys must retry the case.[188] However, mistrial orders come with a risk to the government.[189] Under some circumstances, a retrial may be barred under constitutional double jeopardy.[190] Thus, prosecutors seeking to penalize attorney misconduct via mistrial must proceed with care.

v. *28 U.S.C. § 1927*

Another tool the court could use to curb the use of the deepfake defense, also employed in *King v. Witmer*, is 28 U.S.C. § 1927.[191] Under Section 1927, any attorney "who so multiplies the proceedings in any case unreasonably and vexatiously may be required by the court to satisfy personally the excess costs, expenses, and attorneys' fees reasonably incurred because of such conduct."[192] The purpose of a sanctions award under this provision is to "deter dilatory litigation practices and to punish aggressive tactics that far exceed zealous advocacy."[193]

---

evidence). Some states adopt variations of Rule 11 for their criminal proceedings. For example, New Mexico adopts essentially the 1938 Rule 11 for their criminal proceedings. N.M. R. CRIM. P. DIST. CTS. 5-206 (1997). Similarly, Texas's signature rule for criminal cases is the same as its Rule 11 analogue for civil cases. TEX. CODE CRIM. PROC. ANN. art. 1.052 (West 1997).

[186] Arizona v. Washington, 434 U.S. 497, 497 (1978).

[187] *See Arizona*, 434 U.S. at 500–01; *see also* FED. R. CRIM. P. 26.3.

[188] *See* United States v. Wecht, 541 F.3d 493, 499 (3d Cir. 2008) (explaining that a retrial is permitted when the first trial ended with a properly declared mistrial).

[189] 21 AM. JURIS. 2D *Criminal Law* § 337 (2023) (discussing the risk of the Double Jeopardy Clause barring a second trial).

[190] *See id.*; *Arizona*, 434 U.S. at 505. Retrial after mistrial avoids double jeopardy only if the mistrial is of "manifest necessity" under the circumstances. *Id.* at 505–06. In *Arizona*, the Supreme Court held that double jeopardy does not bar retrial when defense counsel's statements likely irreparably prejudice a jury against the government. *Id.* at 515–16. But if mistrial is not manifestly necessary, then the ultimate result of a mistrial will be preclusion of retrying the case.

[191] King v. Whitmer, 556 F. Supp. 3d 680, 696 (E.D. Mich. 2021).

[192] 28 U.S.C. § 1927; Ridder v. City of Springfield, 109 F.3d 288, 298 (6th Cir. 1997). For a general discussion of 28 U.S.C. § 1927, see Janet Eve Josselyn, Note, *The Song of the Sirens—Sanctioning Lawyers Under 28 U.S.C. § 1927*, 31 B.C. L. REV. 477, 479 (1990).

[193] Red Carpet Studios Div. of Source Advantage, Ltd. v. Sater, 465 F.3d 642, 646 (6th Cir. 2006).

Section 1927 imposes an objective standard of conduct on attorneys, and courts need not make a finding of subjective bad faith before assessing monetary sanctions.[194] Instead, a court need only determine that "an attorney . . . reasonably should know that a claim pursued is frivolous."[195] "Simple inadvertence or negligence, however, will not support sanctions under § 1927."[196] Ultimately, "[t]here must be some conduct on the part of the subject attorney that trial judges, applying collective wisdom of their experience on the bench, could agree falls short of the obligations owed by a member of the bar to the court . . . ."[197]

The statute applies to all proceedings—civil and criminal—in federal courts of all levels.[198] It could punish false arguments and statements that embody the deepfake defense.[199]

vi. *The Procedural Rules and Statutes Are Inadequate*

Applying the existing rules and law governing procedure, Rules 11, 59, 60, and 28 U.S.C. § 1927, will not solve the problem of the deepfake defense. For example, Rule 11 may prove ineffective in addressing the challenge of lawyer misconduct in connection with the deepfake defense. First, Rule 11's application is limited to civil proceedings; thus, lawyers in criminal cases will not be constrained by any threat of sanctions that Rule 11 presents. Second, Rule 11, as described elsewhere, only curbs litigation conduct connected to papers and documents "filed in court by an attorney, not to questionable attorney conduct in general."[200] And it only applies only to signed papers and not to oral statements.[201] A trial attorney's effort to exploit doubts about the authenticity

---

[194] *See id.* (citing Jones v. Cont'l Corp., 789 F.2d 1225, 1230 (6th Cir. 1986)).

[195] Jones v. Cont'l Corp., 789 F.2d 1225, 1230 (6th Cir. 1986).

[196] Salkil v. Mount Sterling Twp. Police Dep't, 458 F.3d 520, 532 (6th Cir. 2006) (citing *Ridder*, 109 F.3d at 298); *see also Red Carpet Studios*, 465 F.3d at 646 (holding that "§ 1927 sanctions require a showing of something less than subjective bad faith, but something more than negligence or incompetence").

[197] *Ridder*, 109 F.3d at 298 (quoting *In re* Ruben, 825 F.2d 977, 984 (6th Cir. 1987)).

[198] *See, e.g.*, *In re* Ginther, 791 F.2d 1151, 1156 (5th Cir. 1986) (civil case); Wisconsin v. Glick, 782 F.2d 670, 673 (7th Cir. 1986) (criminal case); *In re* Usoskin, 56 B.R. 805, 819 (Bankr. E.D.N.Y. 1985) (bankruptcy proceeding).

[199] Escribano-Reyes v. Pro. Hepa Certificate Corp., 817 F.3d 380, 391 (1st Cir. 2016) (the submission of false statement to the court supported the imposition of sanctions under section 1927); *see also* Haeger v. Goodyear Tire and Rubber Co., 906 F. Supp. 2d 938, 938 (D. Ariz. 2012), *aff'd*, 793 F.3d 1122 (9th Cir. 2015), *and* 813 F.3d 1233 (9th Cir. 2016), *vacated and remanded*, 869 F.3d 707 (9th Cir. 2017) (sanctions appropriate based on counsel's misrepresentations to the court).

[200] Jackson v. Law Firm of O'Hara, Ruberg, Osborne & Taylor, 875 F.2d 1224, 1229 (6th Cir. 1989).

[201] Atl. Mut. Ins. Co. v. Balfour Maclaine Int'l Ltd., 775 F. Supp. 101, 110 (S.D.N.Y.1991), *aff'd* 968 F.2d 196 (N.Y. 1992). Although 1993 version of Rule 11 permits imposition of sanctions based on litigants' oral representations, not all oral statements are

of audiovisual evidence may not be memorialized in a document filed in court. Instead, it may materialize in closing arguments to the jury. Thus, the counsel's effort to cash in on the deepfake defense is likely beyond the reach of Rule 11. And even in circumstances when Rule 11 might apply, the court may not use it. Historically, Rule 11 sanctions have been sparingly imposed to avoid chilling or stifling advocacy.[202]

Obtaining relief for counsel's use of the deepfake defense under Rule 59 is equally unlikely. New trials because of attorney misconduct are rare.[203] One commentator described it as an elusive and ethereal remedy.[204] In general, courts only order new trials where counsel engaged in multiple instances of serious misconduct in a trial.[205] Nachman writes that "[c]ourts hold that conduct of counsel is ordinarily not grounds for reversal, unless such conduct substantially influences the verdict or denies a party a fair trial."[206] However, assessing the potential effect of attorney misconduct on the jury will be speculative. "Neither the court nor the parties are allowed to question jurors about how they reached their verdict, no matter how interesting and useful that exercise might be."[207] Some courts forbid counsel from communicating with jurors unless permitted by the court or if initiated by a juror.[208] "The safest prediction is that granting a new trial for attorney misconduct is the exception rather than the rule, and it is probable that the appeals court will affirm the trial court's ruling, whether favorable or not."[209]

Likewise, the ability of a court to vacate a judgment based on fraud between the parties or upon the court under Rule 60(b) might potentially provide a remedy for the victims of misrepresentations occurring during the litigation process. However, parties seeking relief under Rule 60(b)(3) face several potential obstacles. First, courts often require that a party prove fraud with clear and convincing evidence, which is a high burden.[210] In addition, the moving party must establish that the fraud prevented the party from "fully and fairly"

---

sanctionable under Rule 11, even when they advance baseless allegations or objectively frivolous arguments. O'Brien v. Alexander, 101 F.3d 1479, 1479 (2d Cir. 1996).

[202] *See, e.g.*, Guzzello v. Venteau, 789 F. Supp. 112, 118 (E.D.N.Y.1992).

[203] Franklin A. Nachman, *When the Sideshow Consumes the Circus: New Trials on Account of Attorney Misconduct*, 36 LITIG. 31, 31 (2010) (surveying the case interpreting and applying Rule 59 to order new trial for a range of jury misconduct).

[204] Nachman observed the "bright-line rules for granting new trials on account of attorney misconduct under Rule 59(a)(1)(A) of the Federal Rules of Civil Procedure may be compared to the yellow first down marker on a football field. The precedents, like the yellow marker, seem real when you first read them, but like the marker, they tend to disappear when it is time to decide whether a new trial motion will succeed." *Id.*

[205] *Id.* at 32.

[206] *Id.* at 31.

[207] *Id.*; *see also* FED. R. EVID. 606(b).

[208] Nachman, *supra* note 203, at 31.

[209] *Id.*

[210] 11 CHARLES ALAN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE § 2860 (3d ed. 2023).

presenting their case.[211] Parties seeking relief under Rule 60(d)(3) based upon fraud upon the court face obstacles as well. Courts tend to vacate judgments on this basis only where there is "the most egregious conduct involving a corruption of the judicial process itself," such as bribery of a judge.[212] For example, according to Wright and Miller, there are only "a few cases" that treat perjury as the type of fraud upon the court that warrants vacating a judgment.[213] However, the fact that an attorney was involved in the fraud may be a relevant consideration.[214]

Although 28 U.S.C. § 1927 offers one possible solution to counter the deepfake defense because of its broad application to criminal and civil proceedings, and unlike Rule 11 does not depend on signed writing, its contours are not clearly defined.[215] Courts do not agree on the type of conduct that constitutes an "unreasonable and vexatious" multiplication of the proceedings.[216] Whether an attorney has increased the costs "unreasonably and vexatiously" within the meaning of section 1927 depends upon whether the court determines that the phrase "unreasonable and vexatious" implies a bad faith or intentional misconduct requirement not explicit in the statutory language.[217] A review of the federal judicial circuits reveals that there is currently no uniform standard against which an attorney's conduct is measured to determine whether an attorney has multiplied the proceedings "unreasonably and vexatiously."[218]

---

[211] *Id.*

[212] *Id.* § 2870; Roger Edwards, LLC v. Fiddes & Son Ltd., 427 F.3d 129, 133 (1st Cir. 2005) (stating that conduct in question must be severe and stating that "perjury alone . . . has never been sufficient" (quotations omitted)); *see also* Toscano v. Comm'r, 441 F.2d 930, 933–34 (9th Cir. 1971) (stating that term "fraud upon the court" must be construed narrowly in connection with Rule 60).

[213] 11 WRIGHT & MILLER, *supra* note 210, § 2870.

[214] *Id.*

[215] *See, e.g.*, Josselyn, *supra* note 192, at 481 (describing the case law interpreting section 1927 and lamenting that the courts do not apply the standard for assessing sanctionable conduct under the statute in a clear and consistent manner).

[216] *Id.*

[217] Barnd v. City of Tacoma, 664 F.2d 1339, 1343 (9th Cir. 1982).

[218] In United States v. Nesglo, Inc., 744 F.2d 887, 892 (1st Cir. 1984) and *In re* Oximetrix, Inc., 748 F.2d 637, 644 (Fed. Cir. 1984), the courts affirmed the district courts' award of section 1927 sanctions because bad faith was clearly present. In both cases, however, the courts failed to indicate whether bad faith was or was not a requirement for the imposition of section 1927 sanctions. In Limerick v. Greenwald, 749 F.2d 97, 102 (1st Cir. 1984), the court affirmed the section 1927 sanction without adequate indication as to the standard the court employed. In O'Connell v. Champion Int'l Corp., 812 F.2d 393, 395 n.2 (8th Cir. 1987), the court stated that the question of whether section 1927 required a finding of bad faith in addition to unreasonable conduct was not before the court. Similarly, in Hashemi v. Campaigner Publ'ns, Inc., 784 F.2d 1581, 1584 (11th Cir. 1986) and Amey, Inc. v. Gulf Abstract & Title, Inc., 758 F.2d 1486, 1510 (11th Cir. 1985), both courts of appeals held that the district courts did not abuse their discretion in failing to award or in denying motions to award sanctions pursuant to section 1927. The courts, however, failed to indicate

c. *The Substantive Law and Its Limitations*

In addition to the inherent power of the courts and the procedural rules and statutes to address attorney use of the deepfake defense, the substantive law must be considered as a possible solution to the use of the deepfake defense. This section examines the possible criminal and civil remedies that courts and parties might look to address an attorney's misconduct during proceedings or trials and why they do not provide an adequate remedy.

i. *Criminal Actions*

Given the brazenness of the misconduct in the use of the deepfake defense and the significant harms that it sows both for the individual and the justice system, the first source that one might consider are criminal penalties. Although the general criminal law applies equally to lawyers and nonlawyers,[219] there are only a few criminal statutes—such as those prohibiting barratry, which is the act of encouraging lawsuits between others to create legal business or the hiring of runners to solicit employment—that specifically single out lawyers for punishment.[220] One such crime is attorney deceit. More than a dozen jurisdictions—including California and New York—have statutes that single out lawyers who engage in deceit or collusion.[221] In nearly all these

---

their reasons. In Blair v. Shenandoah Women's Ctr., Inc., 757 F.2d 1435, 1438 (4th Cir. 1985), it is unclear what authority or statute the court relied on for the imposition of sanctions.

[219] *See* RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 8 cmt. c (AM. L. INST. 2000) ("For the most part, the substantive law of crimes applicable to lawyers is that applicable to others."); Fred C. Zacharias, *Integrity Ethics*, 22 GEO. J. LEGAL ETHICS 541, 559 (2009) (noting that "lawyers are subject to criminal law and that nothing about the roles prescribed in [ethics] codes excuses lawyers from abiding by laws of general applicability"); Bruce A. Green, *The Criminal Regulation of Lawyers*, 67 FORDHAM L. REV. 327, 330–31 (1998) (suggesting that existing scholarship underestimates extent to which criminal law regulates lawyers' conduct).

[220] *See, e.g.*, CAL. BUS. & PROF. CODE § 6152 (West 2003) (prohibiting lawyers from using "runner[s]" to solicit employment); TEX. PENAL CODE ANN. § 38.12(a)(3) (West 2003) (prohibiting barratry).

[221] *See, e.g.*, CAL. BUS. & PROF. CODE § 6128 (West 1976); IND. CODE ANN. § 33-43-1-8 (West 2023); IOWA CODE ANN. § 602.10113 (West 2023); MINN. STAT. ANN. § 481.07 (West 1986); MINN. STAT. ANN. § 481.071 (West 1986); MONT. CODE ANN. § 37-61-406 (West 2009); NEB. REV. STAT. ANN. § 7-106 (1963); N.M. STAT. ANN. § 36-2-17 (West 1978); N.Y. JUD. LAW § 487 (McKinney 2014); N.D. CENT. CODE ANN. § 27-13-08 (West 1975); OKLA. STAT. ANN. tit. 21, § 575 (West 1910); S.D. CODIFIED LAWS § 16-18-26 (1979); WYO. STAT. ANN. § 33-5-114 (West 2015). Utah's statute has been repealed. Bennett v. Jones, Waldo, Holbrook & McDonough, 70 P.3d 17, 33 (Utah 2003). New York's statute, section 487 of the Judiciary Law, is typical: "An attorney or counselor who: 1. Is guilty of any deceit or collusion, or consents to any deceit or collusion, with intent to deceive the court or any party; or, 2. Willfully delays his client's suit with a view to his own gain; or, willfully receives any money or allowance for or on account of any money which he has not laid out,

jurisdictions, a lawyer found to have engaged in such action faces criminal penalties or civil liability, or both.[222] Most jurisdictions designate such conduct as a misdemeanor and allow an injured party to recover treble damages in a civil action.[223] A few provide for recovery of treble damages by an injured party and mention the possibility of disbarment for the offending attorney but do not criminalize the deceit or collusion.[224] Two designate such conduct as a misdemeanor but make no mention of disbarment or treble damages.[225] One (Nebraska) advises that a lawyer who engages in such conduct is subject to disbarment.[226]

While the criminal deceit statutes might be available to reign in lawyer misconduct, convincing a prosecutor to file such a case would likely prove challenging. And those dozen states with such statutes apply them infrequently and restrictively.[227] The fact that most people have paid little attention to these attorney deceit statutes is understandable. Until recently, the statutes have languished in obscurity, and courts have rendered them somewhat irrelevant through a series of restrictive readings of the statutory language.[228]

ii. *Civil Actions*

In theory, a party might consider asserting an intentional tort claim against an attorney who has engaged in deceptive conduct, including the reliance on the deepfake defense in legal proceedings. The traditional and well-established tort claims, such as fraud and defamation, and less common claims, such as

---

or becomes answerable for, Is guilty of a misdemeanor, and in addition to the punishment prescribed therefor by the penal law, he forfeits to the party injured treble damages, to be recovered in a civil action." N.Y. JUD. LAW § 487; *see* IND. CODE ANN. § 33-43-1-8. Some jurisdictions do not specifically prohibit an attorney from willfully delaying a client's suit with a view to the lawyer's own gain as does New York's Judiciary Law § 487(2). IND. CODE ANN. § 33-43-1-8; IOWA CODE ANN. § 602.10113; MONT. CODE ANN. § 37-61-406; N.M. STAT. ANN. § 36-2-17. Every statute, however, prohibits an attorney from engaging in deceit or collusion, or consenting thereto, with the intent to deceive.

[222] Amalfitano v. Rosenberg, 903 N.E.2d 265, 267 (N.Y. 2009) (noting that New York's descends from the first Statute of Westminster and quoting L. 1787, ch. 35, § 5). In addition, the statute provided that the guilty attorney would also pay the costs of suit. *Id.*

[223] *See* IND. CODE ANN. § 33-43-1-8; MINN. STAT. ANN. § 481.07; MONT. CODE ANN. § 37-61-406; N.Y. JUD. LAW § 487; N.D. CENT. CODE ANN. § 27-13-08; OKLA. STAT. ANN. tit. 21, § 575 (2002); CAL. PENAL CODE § 160 (Deering 1915).

[224] *See* IOWA CODE ANN. § 602.10113; N.M. STAT. ANN. § 36-2-17; WYO. STAT. ANN. § 33-5-114.

[225] CAL. BUS. & PROF. CODE § 6128; S.D. CODIFIED LAWS § 16-18-26.

[226] NEB. REV. STAT. ANN. § 7-106.

[227] *See* Alex B. Long, *Attorney Deceit Statutes: Promoting Professionalism Through Criminal Prosecutions and Treble Damages*, 44 U.C. DAVIS L. REV. 413, 450–52 (2010) (describing the myriad and generally narrow interpretations jurisdictions have applied to criminal deceit statutes).

[228] *See generally id.* (arguing for a re-examination and broader application of criminal deceit statutes directed at lawyer misconduct).

malicious defense, are briefly described here. However, as this section explains, those looking to deter the use of the deepfake defense by looking to substantive civil tort claims will not get far.

d. *Defamation and Fraud*

The tort of defamation is an option for a party defamed by an attorney who asserts false allegations regarding a visual image that caused harm. Indeed, the significant harm of deepfakes is that they imply something false about the subject of the deepfake, and challenging the genuineness of an image would likewise impose similar harm.[229] Asserting a fraud claim might be another theoretical possibility. An attorney who intentionally (or negligently) presents a false statement to challenge the genuineness of a digital image, which the lawyer knows is real, may harm their adversary.[230] However, a fraud claim would likely fail because to prevail on a fraudulent misrepresentation claim; a plaintiff must establish not only that the defendant made a false statement of fact but that the plaintiff justifiably relied on the misrepresentation to their detriment.[231] If anyone is likely to be deceived by false allegations in a court filing, it is the trier of fact, not the victim of the false allegations. Thus, the fraud is perpetrated (if at all) upon the court, not the subject of the misrepresentation.[232]

Several other significant limitations also exist to asserting fraud or defamation against counsel who uses the deepfake defense. The litigation privilege is one major impediment to asserting a tort claim against a lawyer who used the deepfake defense.[233] According to Section 586 of the Restatement (Second) of Torts, "[a]n attorney . . . is absolutely privileged to publish defamatory matter concerning another in communications preliminary to a proposed judicial proceeding, or in the institution of, or during the course and

---

[229] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 299–302 (discussing the harms from deepfakes).

[230] *See* Long, *supra* note 227, at 450.

[231] *See id.*; RESTATEMENT (SECOND) OF TORTS § 552(1) (AM. L. INST. 1977).

[232] *See generally* Amalfitano v. Rosenberg, 533 F.3d 117, 119, 124 (2d Cir. 2008) (classifying misrepresentations in complaint and in support of summary judgment motion as deceit upon court). The same justification has been offered for the general prohibition against seeking civil damages for perjury. *See* Kessler v. Townsley, 182 So. 232, 232–33 (Fla. 1938); RESTATEMENT (SECOND) OF TORTS § 588 (AM. L. INST. 1977). Nearly every jurisdiction has concluded that there is no civil cause of action for perjury. *See* Cooper v. Parker-Hughey, 894 P.2d 1096, 1100–01 (Okla. 1995) (listing Maine as only jurisdiction to recognize such action and citing cases). Because a witness's false statements amount to a fraud upon the court or jury, rather than a litigant, there is no reliance on the part of the litigant; thus, a common law fraudulent misrepresentation claim would not cover perjurious testimony. *Id.* at 1100.

[233] This privilege is also referred to as the "judicial proceedings privilege," the "judicial privilege," or the "defamation privilege." *See* Messina v. Krakower, 439 F.3d 755, 760 (D.C. Cir. 2006); Buchanan v. Minn. State Dep't of Health, 573 N.W.2d 733, 736 (Minn. Ct. App. 1998); Bochetto v. Gibson, 860 A.2d 67, 71 (Pa. 2004).

as a part of, a judicial proceeding in which he participates as counsel if it has some relation to the proceeding."[234] The privilege provides lawyers with absolute immunity from civil liability for statements related to the litigation.[235] The fact that a lawyer has good reason to suspect or has actual knowledge that their statement to the court is untrue does not deprive the lawyer of the privilege.[236] The privilege would arguably extend to the counsel's use of the deepfake defense.[237] The privilege provides attorneys the "utmost freedom in their efforts to secure justice for their clients."[238] In addition, the privilege is justified based on the rationales that the integrity of the adversarial system outweighs any financial harm imposed on the injured party[239] and that other remedies to curb lawyer misconduct exist, including procedural rules, the court's inherent power to sanction lawyers and bar disciplinary proceedings.

Another limitation to using tort remedies to curb the deepfake defense is the widely accepted view that a lawyer owes no duty of care to an opposing party.[240] Thus, absent unusual circumstances, a lawyer who makes a false statement of material fact to the opposing side does not face liability under a fraud theory.[241] Although the "no-duty rule" has been most commonly expressed in the context of claims of negligence, courts have cited the rule in shielding lawyers from liability where the lawyers have been accused of fraud resulting from the failure to disclose material information, including the failure to disclose the fact that the lawyer's client has made a fraudulent statement.[242]

e. *Malicious Defense*

The tort of malicious defense is another potential legal mechanism to hold the lawyers to account for knowingly making false assertions about genuine

---

[234] RESTATEMENT (SECOND) OF TORTS § 586 (AM. L. INST. 1977).

[235] Louise Lark Hill, *The Litigation Privilege: Its Place in Contemporary Jurisprudence*, 44 HOFTSRA L. REV. 401, 401 (2015).

[236] RESTATEMENT (SECOND) OF TORTS § 586 cmt. a (AM. L. INST. 1977) (explaining scope of privilege).

[237] *See, e.g.*, Lark Hill, *supra* note 235, at 402–03 (litigation privilege can protect attorneys regardless of motive).

[238] RESTATEMENT (SECOND) OF TORTS § 586 cmt. a (AM. L. INST. 1977).

[239] Lark Hill, *supra* note 235, at 401–02.

[240] RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 51 cmt. c (2000); *see, e.g.*, Hugel v. Milberg, Weiss, Bershad, Hynes & Lerach, LLP, 175 F.3d 14, 17–18 (1st Cir. 1999) (holding an attorney has duties to her client, she has no duties to other parties); Garcia v. Rodey, Dickason, Sloan, Akin & Robb, 750 P.2d 118, 122 (N.M. 1988).

[241] *See, e.g.*, *Garcia*, 750 P.2d at 122 (asserting that "[n]egligence is not a standard on which to base liability of an attorney to an adverse party").

[242] *See, e.g.*, Schatz v. Rosenberg, 943 F.2d 485, 492 (4th Cir. 1991) (rejecting fraud claim against attorney based on failure to disclose client's misrepresentations); Schlaifer Nance & Co. v. Est. of Warhol, 927 F. Supp. 650, 661 (S.D.N.Y. 1996) (dismissing fraud claims against attorney based on attorney's failure to volunteer information and failure to correct client's false statement).

audio-visual evidence.[243] The tort, recognized in New Hampshire, applies to one who initiates or continues a defense in a civil proceeding without probable cause primarily for an improper purpose (such as to delay or harass) and who causes damages may be liable under a malicious defense theory.[244] Damages in this context include emotional distress and the expense incurred in defending oneself in the proceeding.[245] In theory, the tort could be broad enough to cover a variety of litigation tactics, including false assertions or introducing fabricated evidence in support of a motion or during an argument to the court or the jury.[246]

However, most jurisdictions do not recognize this tort because the availability of judicial sanctions for "frivolous or delaying conduct" is an adequate deterrent to such misconduct.[247] Also, some might argue that permitting such claims "may 'have a chilling effect on some legitimate defense and perhaps drive a wedge between defendants seeking zealous advocacy and defense attorneys who fear personal liability in a second action'" and, relatedly, that permitting such claims would threaten the absolute litigator's privilege.[248]

---

[243] *See* Jonathan K. Van Patten & Robert E. Willard, *The Limits of Advocacy: A Proposal for the Tort of Malicious Defense in Civil Litigation*, 35 HASTINGS L.J. 891, 894 (1984) (noting conceptual similarity of claims and stating that failure to proscribe malicious defense encourages dishonesty).

[244] Aranson v. Schroeder, 671 A.2d 1023, 1028–29 (N.H. 1995). As stated by the New Hampshire Supreme Court:

> One who takes an active part in the initiation, continuation, or procurement of the defense of a civil proceeding is subject to liability for all harm proximately caused, including reasonable attorneys' fees, if
> (a) he or she acts without probable cause, i.e., without any credible basis in fact and such action is not warranted by existing law or established equitable principles or a good faith argument for the extension, modification, or reversal of existing law,
> (b) with knowledge or notice of the lack of merit in such actions,
> (c) primarily for a purpose other than that of securing the proper adjudication of the claim and defense thereto, such as to harass, annoy or injure, or to cause unnecessary delay or needless increase in the cost of litigation,
> (d) the previous proceedings are terminated in favor of the party bringing the malicious defense action, and
> (e) injury or damage is sustained.

Id. at 1028–29.

[245] *See id.* at 1028.

[246] *See id.* at 1025.

[247] Sheldon Appel Co. v. Albert & Oliker, 765 P.2d 498, 503 (Cal. 1989).

[248] Young v. Allstate Ins. Co., 198 P.3d 666, 682–83, 684 (Haw. 2008).

2. *Ethical Rules and Their Limitations*

An attorney's use of the deepfake defense raises ethical concerns. In law practice, honesty and integrity are not simply a matter of personal morality or values. They are deemed essential to lawyers' role because the effectiveness and efficiency of most aspects of law practice depend on others—judges, clients, colleagues, and other lawyers—being able to trust lawyers and take them at their word.[249] This understanding pervaded nineteenth-century writings on the legal profession,[250] and it was incorporated in the ABA's Canons of Professional Ethics, which in 1908 became the first national codification of lawyers' professional expectations.[251] It has been a consistent theme of ethics codes and other professional writings since then.[252] In 1983, the American Bar Association's House of Delegates adopted the Model Rules of Professional Conduct (hereafter the "Model Rules").[253] Most jurisdictions have adopted some version of model rules.[254] The current rules address a range of professional conduct that one might regard as dishonest, deceitful, or lacking in candor, including, in various contexts, the failure to correct false statements,[255] and

---

[249] *See, e.g.*, State *ex rel.* Kilbourn v. Hand, 9 Ohio 42, 42 (1839) ("The discharge of professional duties, demands great and unreserved confidence from the client, and the connection of the attorney with courts, and his access to papers, require unsuspected integrity. Hence general honesty and fidelity to clients, is not only necessary to his success, but even to the performance of his duties. Other good qualities may be wanting in his character, and some vices may be present, but these are the essential virtues of his calling, no more to be dispensed with than courage in a soldier, or modesty in a woman." (emphasis in original)).

[250] *Id.*

[251] *See, e.g.*, CODE OF PRO. ETHICS Canon 32 (AM. BAR ASS'N 1908) ("[A]bove all a lawyer will find his highest honor in a deserved reputation for fidelity to private trust and to public duty, as an honest man and as a patriotic and loyal citizen.").

[252] *See, e.g.*, MODEL CODE OF PRO. RESP. DR 1-102(A)(4) (AM. BAR ASS'N 1980) (subjecting a lawyer to professional discipline for "conduct involving dishonesty, fraud, deceit, or misrepresentation"); Charles P. Curtis, *The Ethics of Advocacy*, 4 STAN. L. REV. 3, 3 (1951); Alvin B. Rubin, *A Causerie on Lawyers' Ethics in Negotiation*, 35 LA. L. REV. 577, 580 (1975).

[253] *Model Rules of Professional Conduct*, AM. BAR ASS'N, https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/ [https://perma.cc/56hv-c4x2].

[254] *See Additional Legal Ethics and Professional Responsibility Resources*, AM. BAR ASS'N, https://www.americanbar.org/groups/professional_responsibility/resources/links_of_interest/#States [https://perma.cc/C5UX-T69T] (providing links to States' codes of professional responsibility).

[255] MODEL RULES OF PRO. CONDUCT r. 3.3(a) (AM. BAR ASS'N 2021) ("A lawyer shall not knowingly . . . fail to correct a false statement of material fact or law previously made to the tribunal by the lawyer.").

other nondisclosures;[256] recklessly false statements;[257] statements that are misleading though not necessarily literally false;[258] and misleading conduct.[259]

Of course, lawyers' professional obligations are contextual.[260] For example, concerning lawyers' candor and truthfulness, the current rules distinguish between trial advocacy (or the equivalent) and lawyers' many other pursuits. But in advocacy and other legal work, certain statements that the public would regard as lies are conceived as something other than lies or otherwise permitted. As a matter of policy, courts make judgments distinguishing between lies that are bad and, therefore, sanctionable and those that are good or at least innocuous and, therefore, permissible. This part describes the ethical rules that are implicated by the deepfake defense.

a. *ABA Model Rules That Might Apply to the Deepfake Defense*

The professional conduct rules include several provisions targeting false statements and other false or deceitful conduct in the professional setting that might apply to the deepfake defense.

i. *Model Rule 3.1: Meritorious Claims & Contentions*

Part 3 of the Model Rules governs the conduct of an attorney acting as an advocate on a client's behalf.[261] Rule 3.1 requires that attorneys only assert claims with a basis in law and fact so "that is not frivolous, which includes a good faith argument for an extension, modification or reversal of existing law."[262]

Lawyers have been sanctioned for violating this rule by presenting *frivolous* allegations to the court.[263] The issue in most Rule 3.1 cases is straightforward:

---

[256] *Id.* r. 3.3(a)(2) (duty to disclose certain adverse legal authority); *id.* r. 3.3(d) (duty to disclose material facts in an ex parte proceeding).

[257] MODEL RULES OF PRO. CONDUCT r. 8.2(a) (AM. BAR ASS'N 1983) (forbidding making a false statement about a judge's integrity or qualifications either knowingly "or with reckless disregard as to its truth or falsity").

[258] MODEL RULES OF PRO. CONDUCT r. 4.1 cmt. 1 (AM. BAR ASS'N 2021) (misrepresentations in violation of Rule 4.1(a) "can also occur by partially true but misleading statements").

[259] *Id.* r. 8.4(c) (forbidding conduct involving dishonesty or deceit).

[260] *See, e.g.*, Bruce A. Green, *Less Is More: Teaching Legal Ethics in Context*, 39 WM. & MARY L. REV. 357, 382–85 (1998) (describing "the importance of context in resolving ethical dilemmas and in defining the lawyer's role and responsibilities").

[261] MODEL RULES OF PRO. CONDUCT r. 3 (AM. BAR ASS'N 2021).

[262] *Id.* r. 3.1.

[263] *See, e.g.*, Brunswick v. Statewide Grievance Comm., 931 A.2d 319, 323–24 (Conn. App. 2007); *In re* Disciplinary of Shea, 273 P.3d 612, 621 (Alaska 2012) (concluding that a defense attorney violated Rule 3.1 when he accused the plaintiff's attorney of criminal conduct and provided no factual basis to support his assertions).

the attorneys failed to present evidence for their client's claims.[264] On the other hand, a deepfake defense is meant to undermine reality, to make people question the validity of the legitimate.[265] A lawyer using a deepfake defense can argue that they have informed themselves of the facts and that their facts show that this video, this photo, or this recording is fake. It is much easier to identify a Rule 3.1 violation when the evidence does not exist. It is more difficult to find a Rule 3.1 violation when the prosecution's, or plaintiff's, evidence *exists,* and the defense asserts that it is fake.

Lastly, the generality of Rule 3.1, like the other ethical rules, results in inconsistent sanctions. While every case is unique, sanctions tend to run the gamut from reprimanded to suspension for 25 months.[266] In addition, courts will often look at how judges sanctioned similar cases in their jurisdiction and use that as a guideline.[267] This leads to inconsistent results across jurisdictions for the same rule violations.

ii. *Model Rule 3.3: Candor Toward the Tribunal*

The Model Rules make a powerful statement regarding lawyers' honesty in the context of courtroom advocacy. Model Rule 3.3(a) states, "[a] lawyer shall not knowingly . . . make a false statement of fact or law to a tribunal."[268] Underscoring the importance of "candor toward the tribunal," the accompanying Comment interprets the phrase "false statement" broadly insofar as it recognizes that the rule may cover misleading silence and affirmative false statements.[269]

The Comment does not characterize Rule 3.3 as a particular application of a general duty to be truthful but explains that the rule arises out of advocates' special relationship to the courts, in that the rule expresses "the special duties of lawyers as officers of the court to avoid conduct that undermines the integrity of the adjudicative process."[270] The implication is that lawyers' obligations in communications with the courts are especially stringent. Rule 3.3, in short, is used to sanction lawyers who lie or lie by omission.[271] Thus, an attorney using a deepfake defense may violate Rule 3.3 by making a false statement of fact. However, the court must find that the lawyer did so *knowingly*. Conversely, it is difficult to prove that an attorney asserting a deepfake defense is knowingly

---

[264] As the court in the first case stated, "[t]he allegation of fraud, corruption or undue influence . . . was clearly frivolous, as the [plaintiff] had no evidence to support the allegation." *Brunswick*, 931 A.2d at 323.

[265] Delfino, *Deepfakes on Trial*, *supra* note 13, at 310.

[266] *Brunswick*, 931 A.2d at 322; *In re* Disciplinary of Shea, 273 P.3d at 623.

[267] *In re* Rios, 109 A.D.3d 64, 70–71 (N.Y. App. Div. 2013).

[268] MODEL RULES OF PRO. CONDUCT r. 3.3 (AM. BAR ASS'N 2021).

[269] *See id.* cmt. 3 ("There are circumstances where failure to make a disclosure is the equivalent of an affirmative misrepresentation.").

[270] *Id.* r. 3.3 cmt. 2.

[271] *See id.*

espousing a lie when the only tangible evidence to disprove the defense is the evidence that the attorney is disputing.

Although Rule 3.3(a) conveys that candor is at a premium when lawyers speak to judges, several categories of speech are, or may be, excluded. At least in courtroom advocacy, arguments appear to drop off the list of false factual representations—i.e., lies—captured by the candor rules. A lawyer's closing arguments to a jury, for example, do not purport to be based on the lawyer's personal knowledge but are based on evidence introduced at trial. Likewise, a lawyer's arguments to the judge on a motion are ordinarily based on evidence. A lawyer may not knowingly make a false "assertion purporting to be on the lawyer's own knowledge, as in an affidavit by the lawyer or in a statement in open court,"[272] but the rules do not clearly forbid a lawyer from knowingly making false arguments—i.e., false factual assertions premised on others' false statements, false evidence, or erroneous inferences. For example, a lawyer may "argue" that an event occurred, based on inferences from the evidence, even though the lawyer knows that the event never happened. Commentators have debated whether lawyers should make false arguments,[273] but the disciplinary rules do not necessarily foreclose this possibility.

For essentially the same reason, Rule 3.3(a) does not apply to the lawyer's false allegations in adjudicative proceedings, including in pleadings that the lawyer prepares and files in court regarding matters that the lawyer does not purport to have personal knowledge of.[274] The rules restrict frivolous pleadings,[275] as do civil procedure rules,[276] but they do not expressly forbid lawyers from conveying false allegations, as distinguished from false representations. Allegations on behalf of a client are essentially previews of arguments that are expected to be made based on the evidence in the future proceeding. They are not regarded as "statements of fact" under the rule, and,

---

[272] MODEL RULES OF PRO. CONDUCT r. 3.3 cmt. (AM. BAR ASS'N 1983); *see, e.g.*, Pearson v. First NH Mortg. Corp., 200 F.3d 30, 38 (1st Cir. 1999).

[273] *See generally* Liebman, *supra* note 185 (reviewing scholarly and professional literature on whether defense lawyers may use a false defense).

[274] The Comment to Rule 3.3 explains that a lawyer is accountable only for statements "purporting to be on the lawyer's own knowledge." MODEL RULES OF PRO. CONDUCT r. 3.3 cmt. (AM. BAR ASS'N 1983). Knowledge is a defined term in the rules. *See id.* r. 1.0. A lawyer can have knowledge of a fact, and of its truth or falsity, that is not based on direct observation. *See id.* ("A person's knowledge may be inferred from circumstances."). Therefore, a lawyer can conceivably make a knowingly false statement regarding a fact about which the lawyer lacks first-hand knowledge, if the lawyer knows from other sources that the statement is false. But if the lawyer does not purport to have personal knowledge, direct or inferential, the lawyer's statements would presumably fall outside the rule. Guesses, predictions, and expressions of faith are among the kinds of statements that presumably do not qualify as statements of fact because they are not expressions of the lawyer's personal knowledge. *See id.*

[275] MODEL RULES OF PRO. CONDUCT r. 3.1 (AM. BAR ASS'N 1983); *id.* r. 3.3 cmt. (citing r. 3.1).

[276] *See, e.g.*, FED. R. CIV. P. 11.

unless the context or the submission otherwise indicates, they are not expressions of the lawyer's personal knowledge or belief.[277] Consequently, lawyers are not expected to believe their own allegations. The rules suggest that lawyers may therefore make allegations believing and perhaps even knowing them to be false if the allegations are not predicated on perjury or false evidence. Couching falsehoods as allegations or arguments may lead others to believe them, since lawyers convey them. To compound the problem, although advocates may not expressly vouch for false allegations and arguments, advocates may make them with feigned conviction, leading listeners to infer or assume that lawyers believe what they are saying.

The reference to "false statement[s] of fact or law" in Rule 3.3(a) implies other carve-outs for rhetoric that is not a "statement" or that states something other than "fact or law."[278] The rules could be read to exclude statements that merely imply false facts since implications are not statements. The rules could also be read to exclude lawyers' false statements of their opinion, intent regarding future conduct, or general state of mind since these are not necessarily what is meant by facts.[279]

iii. *Model Rule 3.4: Fairness to Opposing Party and Counsel*

Another ethical rule a lawyer relying on the deepfake defense may violate is Rule 3.4.[280] Model Rule 3.4(c) states that a lawyer shall not "knowingly

---

[277] MODEL RULES OF PRO. CONDUCT r. 3.3 cmt. (AM. BAR ASS'N 1983) (stating that a lawyer is not required to have personal knowledge of the allegations in a complaint).

[278] *Id.* r. 3.3(a).

[279] In some contexts, false statements of intent regarding future conduct are not regarded as false statements of fact. *See, e.g.*, Matsumura v. Benihana Nat'l Corp., 542 F. Supp. 2d 245, 253 (S.D.N.Y. 2008) ("Though misrepresentations of present or past fact have the potential to create liability for the speaker, '[m]ere unfulfilled promissory statements as to what will be done in the future are not actionable.'" (quoting Brown v. Lockwood, 432 N.Y.S.2d 186, 194 (N.Y. App. Div. 1980))). *But see In re* Hong-Min Jun, 78 N.E.3d 1100, 1100 (Ind. 2017) (observing that it is a federal crime for a visa applicant to make a false statement of intent to leave the country upon expiration of a visa, and sanctioning a lawyer under Rule 1.2(d) for assisting the client's wife in making a false application).

[280] Model Rule 3.4 provides:

> A lawyer shall not: (a) unlawfully obstruct another party' s access to evidence or unlawfully alter, destroy or conceal a document or other material having potential evidentiary value. A lawyer shall not counsel or assist another person to do any such act; (b) falsify evidence, counsel or assist a witness to testify falsely, or offer an inducement to a witness that is prohibited by law; (c) knowingly disobey an obligation under the rules of a tribunal except for an open refusal based on an assertion that no valid obligation exists; (d) in pretrial procedure, make a frivolous discovery request or fail to make reasonably diligent effort to comply with a legally proper discovery request by an opposing party; (e) in trial, allude to any matter that the lawyer does not reasonably

disobey an obligation under the rules of a tribunal,"[281] and the rules of a tribunal include its evidence rules.[282] Offering inadmissible evidence would seem to be knowingly disobeying court rules, as would asking an improper question and withdrawing it if there is an objection.[283] Rule 3.4(e) states that a lawyer shall not "in trial, allude to any matter that the lawyer does not reasonably believe is relevant or that will not be supported by admissible evidence."[284] This means more than just that a lawyer may not mention inadmissible evidence in his opening statement. An attempt to offer it or get it before the jury would also seem to fall within the idea of an allusion to inadmissible evidence.[285] "A lawyer should not attempt to get before the jury evidence which is improper. In all cases in which [a lawyer] has any doubt about the propriety of any disclosures to the jury, a request should be made for leave to approach the bench and obtain a ruling out of the jury's hearing, either by propounding the question and obtaining a ruling or by making an offer of proof."[286]

iv. *Model Rule 4.1: Truthfulness in Statements to Others*

Model Rule 4.1(a), which governs lawyers' truthfulness in communications with others, states that "[i]n the course of representing a client a lawyer shall not knowingly . . . make a false statement of material fact or law to a third person."[287] But, as the materiality limitation in Rule 4.1(a) illustrates, the rules do not reach all conduct that one might regard as deceptive or dishonest, which

---

> believe is relevant or that will not be supported by admissible evidence, assert personal knowledge of facts in issue except when testifying as a witness, or state a personal opinion as to the justness of a cause, the credibility of a witness, the culpability of a civil litigant or the guilt or innocence of an accused; or (f) request a person other than a client to refrain from voluntarily giving relevant information to another party unless: (1) the person is a relative or an employee or other agent of a client; and (2) the lawyer reasonably believes that the person's interests will not be adversely affected by refraining from giving such information.

MODEL RULES OF PRO. CONDUCT r. 3.4 (AM. BAR ASS'N 1983).

[281] *Id.* r. 3.4(c).

[282] *See id.* r. 3.4 cmt.

[283] *See In re* McDonald, 609 N.W.2d 418, 426 (N.D. 2000) (stating that withdrawing improper evidence after a challenge does not cure an ethical violation).

[284] MODEL RULES OF PRO. CONDUCT r. 3.4(e) (AM. BAR ASS'N 1983).

[285] *See, e.g.*, WILLIAM H. FORTUNE, RICHARD H. UNDERWOOD & EDWARD J. IMWINKELRIED, MODERN LITIGATION AND PROFESSIONAL RESPONSIBILITY HANDBOOK 379–81 (1996) (noting that counsel may allude to inadmissible evidence by deliberately asking an improper question or commenting on the court's rulings).

[286] CODE OF TRIAL CONDUCT § 19(g) (AM. COLL. TRIAL LAW. 1972).

[287] MODEL RULES OF PRO. CONDUCT r. 4.1(a) (AM. BAR ASS'N 2021).

might fall within a definition of "lying."[288] For example, the Comment accompanying that rule explains that "[u]nder generally accepted conventions in negotiation . . . [e]stimates of price or value placed on the subject of a transaction and a party's intentions as to an acceptable settlement of a claim" are generally not "statements of material fact."[289] Bruce Green points out:

> [Comment 2] to Rule 4.1 implies that lawyers are free to tell at least three types of lies: those that third parties would not ordinarily believe, those that third parties might believe but on which they are unlikely to act in reliance, and those on which third parties would not be justified in relying even if they might ordinarily do so. One might argue that lawyers' lies about public events in certain media would fall into one or more of these categories since the public would not be justified in believing and acting in reliance on, what commentators, including lawyer-commentators, say in these media.[290]

v. *Model Rule 8.4: Maintaining the Integrity of the Profession*

Rule 8.4 is the catchall ethics rule implicated by every dishonest or deceitful act by an attorney, regardless of whether the attorney acts in a representational or non-representational capacity. To maintain the integrity of the legal profession, Model Rule 8.4 makes it unethical for an attorney to, among other misconduct, "engage in conduct involving dishonesty, fraud, deceit or

---

[288] Some jurisdictions have rejected the materiality limitation. *See, e.g.*, N.Y. RULES OF PRO. CONDUCT r. 4.1 (2009) ("In the course of representing a client, a lawyer shall not knowingly make a false statement of fact or law to a third person."). But immaterial false statements may still fall outside the rule. For example, the Comment accompanying New York's version of Rule 4.1 explains, "Under generally accepted conventions in negotiation, certain types of statements ordinarily are not taken as statements of fact. Estimates of price or value placed on the subject of a transaction and a party's intentions as to an acceptable settlement of a claim are ordinarily in this category." *Id.* r. 4.1 cmt. 2.

[289] MODEL RULES OF PRO. CONDUCT r. 4.1 cmt. (AM. BAR ASS'N 1983). The secondary literature addressing falsehoods in transactional and settlement negotiations is voluminous. *See, e.g.*, James J. Alfini, *Settlement Ethics and Lawyering in ADR Proceedings: A Proposal to Revise Rule 4.1*, 19 N. ILL. U. L. REV. 255, 266 (1999); Rex R. Perschbacher, *Regulating Lawyers' Negotiations*, 27 ARIZ. L. REV. 75 (1985); Geoffrey M. Peters, *The Use of Lies in Negotiation*, 48 OHIO ST. L.J. 1 (1987); Scott R. Peppet, *Can Saints Negotiate? A Brief Introduction to the Problems of Perfect Ethics in Bargaining*, 7 HARV. NEGOT. L. REV. 83, 91–92 (2002).

[290] Green & Roiphe, *supra* note 82, at 74; *cf.* McDougal v. Fox News Network, 489 F. Supp. 3d 174, 183–84 (S.D.N.Y. 2020) (holding that Tucker Carlson's televised assertion that the plaintiff engaged in extortion was not actionable slander because it was rhetorical hyperbole: "[G]iven Mr. Carlson's reputation, any reasonable viewer 'arrive[s] with an appropriate amount of skepticism' about the statements he makes . . . . Whether the Court frames Mr. Carlson's statements as 'exaggeration,' 'non-literal commentary,' or simply bloviating for his audience, the conclusion remains the same—the statements are not actionable").

misrepresentation,"[291] or "engage in conduct that is prejudicial to the administration of justice."[292]

Rule 8.4 has been recently cited as the source of discipline for lawyers' lies and fabrications designed to confuse and distort the truth and undermine democratic institutions. For example, in June of 2021, Rudolph Giuliani was suspended from the practice of law in New York for violating several Rules of Professional Conduct, including 8.4(c), related to his representation of Donald Trump.[293] While Rule 8.4(c) does not explicitly state that a lawyer must act *knowingly*, courts have interpreted the rule as needing to meet this threshold.[294] The court, however, found no dearth of knowledge on the part of Mr. Giuliani. Mr. Giuliani's claims of illegal counting of ballots in Georgia, "illegal aliens" voting in Arizona, dead people voting in Pennsylvania, and fraudulent voting systems, amongst other claims, not only lacked any grounding in fact but were wholly refutable based on actual evidence.[295] The court explained how Mr. Giuliani's conduct violated each ethical rule and discussed his misconduct's effect on the public.[296] Claims of election fraud are now a common talking point, and it is not so surprising that a court would want to quash the issue as it arises. The deepfake defense, however, is not yet commonplace. While an attorney used deepfake defense in the trial of a January 6th insurrectionist,[297] the defense will not be contained solely to this type of trial in the future. A deepfake defense, unlike election fraud, may not always be seen by a court as "uncontroverted misconduct" and an attack on the administration of justice. For this reason, explicit rules are necessary to combat the defense.

b. *The Ethical Rules Do Not Provide an Answer*

Although the rules of ethics should provide a clear and definitive solution to the assertion of the deepfake defense and state bar regulatory authorities should respond accordingly to discipline those who use it, they do not. The Model Rules do not include any specific rules or commentary regarding deepfakes. Although professional conduct rules generally require lawyers to be truthful, and lawyers take an oath to that effect,[298] the rules do not forbid all

---

[291] MODEL RULES OF PRO. CONDUCT r. 8.4(c) (AM. BAR ASS'N 2021).
[292] *Id.* r. 8.4(d).
[293] *In re* Giuliani, 146 N.Y.S.3d 266, 270–71 (App. Div. 2021).
[294] *Id.* at 271.
[295] *Id.* at 279–80.
[296] *Id.* at 279–83.
[297] Verkouteren, *supra* note 6.
[298] *See, e.g.*, *Lawyer's Oath*, STATE BAR OF MICH., https://www.michbar.org/generalinfo/lawyersoath [https://perma.cc/2K2Y-CM7B] ("I do solemnly swear (or affirm): I will support the Constitution of the United States and the Constitution of the State of Michigan . . . I will employ for the purpose of maintaining the causes confided to me such means only as are consistent with truth and honor, and will never seek to mislead the judge or jury by any artifice or false statement of fact or law.").

lying. The professional conduct rules include several provisions targeting false statements and other false or deceitful conduct in the professional setting.[299] However, none of these rules instruct lawyers emphatically and categorically or include any specific rules or commentary regarding offering or challenging audiovisual evidence such as deepfakes.[300] As interpreted and applied, the honesty and candor rules do not include a comprehensive prohibition on lawyers presenting the deepfake defense or lying in general. On the contrary, rule drafters, courts, and other authorities have approved various conduct that one might otherwise regard as lying, often characterizing it differently and more benignly—for example, as allegations, arguments, pretexting, or puffery.[301]

As discussed, the Model Rules, specifically Rule 3.3, offer some general prohibitions against offering evidence that the lawyer knows to be false[302] and permit counsel to refuse to offer a video she reasonably believes is a deepfake.[303] Likewise, Rule 3.3 prohibits attorneys from making false statements of fact or law to a tribunal.[304] But the ethical issues are opaque on challenging the opposing party's evidence. There is no prohibition in the model rules limiting counsel from asserting a challenge to evidence because it is a deepfake.[305] Likewise, there is nothing in the ethical rules preventing a lawyer from relying on the deepfake defense. Attorneys may feel compelled to cultivate skepticism about the authenticity of all audiovisual evidence, and currently, the rules of professional conduct do not prohibit that conduct. Thus, in an age of technological trickery and an era where the trust in governmental and democratic institutions is low, the lack of clarity in the rules of ethics means that the integrity of the process will rely exclusively on the goodwill of lawyers to look beyond the near-term goal of victory in a particular case, or even just a particular motion hearing, and consider the more significant impact that deepfake accusations could have on our legal system. Thus, without an explicit prohibition in the ethical rules, attorneys may exploit the existence of deepfakes

---

[299] *See* discussion *supra* Part III.B.2 (described the rules of ethics implicated by the deepfake defense).

[300] *See id.*

[301] *See* Green & Roiphe, *supra* note 80, at 70–75 (describing the various contexts in which lawyer lies are expected under the rules of professional conduct).

[302] MODEL RULES OF PRO. CONDUCT r. 3.3(a)(3) (AM. BAR ASS'N 2021) ("A lawyer shall not knowingly . . . offer evidence that the lawyer knows to be false."); Nix v. Whiteside, 475 U.S. 157, 166 (1986) ("Although counsel must take all reasonable lawful means to attain the objectives of the client, counsel is precluded from taking steps or in any way assisting the client in presenting false evidence or otherwise violating the law.").

[303] *See* MODEL RULES OF PRO. CONDUCT r. 3.3(a)(3) (AM. BAR ASS'N 2021) ("A lawyer may refuse to offer evidence . . . that the lawyer reasonably believes is false.").

[304] *Id.* r. 3.3(a)(1) (prohibiting a lawyer from making "a false statement of fact or law to a tribunal or fail to correct a false statement of material fact or law previously made to the tribunal by the lawyer").

[305] *But see id.* (prohibiting a lawyer from making "a false statement of fact or law to a tribunal or fail to correct a false statement of material fact or law previously made to the tribunal by the lawyer").

in a way that may inure to the benefit of their client but that may also undermine the jury truth-finding function.[306]

Another shortcoming to relying heavily on the professional disciplinary process to address attorney deceit is that discipline is relatively rare.[307] According to one author, "[r]esearchers agree that sanctioning rates fall well below the level of sanction-worthy acts that lawyers commit in the aggregate."[308] Although fraudulent behavior would seem more likely to catch the attention of disciplinary authorities than other types of rule violations, some critics have questioned the willingness of disciplinary authorities to prosecute litigation-related misconduct.[309]

Thus, unacceptable lawyer lies often go unpunished in the disciplinary system. Ethics codes outline their prohibitions in vague terms and are "notoriously under-enforced."[310] Bar authorities also struggle with politicization, agency capture, and a persistent lack of resources, which further contribute to the underenforcement of ethics codes.[311] According to one study,

---

[306] Theodore F. Claypoole, *AI and Evidence: Let's Start to Worry*, WOMBLE BOND DICKINSON (Nov. 14, 2019), https://www.womblebonddickinson.com/us/insights/blogs/ai-and-evidence-lets-start-worry [https://perma.cc/N5S9-LSZL] (predicting the efforts of lawyers to exploit the existence of deepfakes through the use of experts and argument to cast doubt on real evidence).

[307] *See* Anita Bernstein, *Pitfalls Ahead: A Manifesto for the Training of Lawyers*, 94 CORNELL L. REV. 479, 487 (2009) (referring to "the relatively rare occasion that an errant lawyer receives some form of professional discipline"); Jenny Roberts, *Ignorance Is Effectively Bliss: Collateral Consequences, Silence, and Misinformation in the Guilty-Plea Process*, 95 IOWA L. REV. 119, 154 (2009) ("Although attorneys are bound to conform their behavior to these state codes, the rules in many instances prove only as effective as the strength and likelihood of their enforcement mechanism.").

[308] Bernstein, *supra* note 307, at 487 n.56 (citing Leslie C. Levin, *The Case for Less Secrecy in Lawyer Discipline*, 20 GEO. J. LEGAL ETHICS 1, 1 (2007)).

[309] *See* Arthur F. Greenbaum, *Judicial Reporting of Lawyer Misconduct*, 77 UMKC L. REV. 537, 552 (2009) (suggesting that "disciplinary counsels, with limited resources, do not believe litigation misconduct . . . is an area they need to police more vigorously"); Joy, *supra* note 151, at 812 ("Lawyer disciplinary enforcement rules and standards for imposing sanctions disfavor lawyer discipline for litigation conduct."); Fred C. Zacharias & Bruce A. Green, *The Duty to Avoid Wrongful Convictions: A Thought Experiment in the Regulation of Prosecutors*, 89 B.U. L. REV. 1, 18 (2009) (noting limited resources on part of disciplinary agencies to limit prosecutions).

[310] Benjamin H. Barton, *The ABA, the Rules, and Professionalism: The Mechanics of Self-Defeat and a Call for a Return to the Ethical, Moral, and Practical Approach of the Canons*, 83 N.C. L. REV. 411, 424 (2005).

[311] *See, e.g.*, Fred C. Zacharias, *What Lawyers Do When Nobody's Watching: Legal Advertising as a Case Study of the Impact of Underenforced Professional Rules*, 87 IOWA L. REV. 971, 997 (2002) ("[R]esource constraints prevent disciplinary authorities from fully enforcing all the professional rules[, and they] must choose among violators that come to their attention on the ba- sis of such factors as the severity of the offense, the deterrent effect of prosecution on this and other offenders, the likely cost of prosecution, the nature of the offender, and the effect of enforcement or lack of enforcement on the image of the bar." (citations omitted)).

"[o]nly about five percent of all complaints result in any sanctions against lawyers," and "the sanctions imposed on lawyers are often light and inconsistent."[312] Indeed, "[o]ver 90 percent of complaints are dismissed, only about 2 percent result in public sanctions, and many complainants never even learn the basis of the dismissal, let alone receive an opportunity to challenge it."[313] Another study found that the most frequently unenforced rules are "those requiring lawyers to report misconduct by other lawyers."[314] Thus, critics have questioned the ability of the disciplinary process to serve as a meaningful deterrent to lawyer misconduct.[315] And this criticism is supported by the data. Studies of bar discipline over the last 20 years show that state bar disciplinary bodies tend to defer to the courts regarding sanctioning lawyer misconduct during litigation.[316] But this tendency becomes particularly problematic for a species of misconduct, such as using the deepfake defense where the rules the courts apply, such as Rule 11, do not prohibit the conduct.

The final concern with the rules of ethics as a tool to curb the deepfake defense relates specifically to the conduct of criminal lawyers. Some commentators have suggested that criminal defense attorneys are held to different standards than other attorneys.[317] The suggestion of a different standard for criminal defense attorneys is not defensible under the rules of ethics. There is no general ethical principle permitting attorneys to practice deception, fraud, and trickery to defend someone accused of a crime. Nor does

---

[312] Leslie C. Levin, *The Emperor's Clothes and Other Tales About the Standards for Imposing Lawyer Discipline Sanctions*, 48 AM. U. L. REV. 1, 8–9 (1999).

[313] Deborah L. Rhode, *The Profession and the Public Interest*, 54 STAN. L. REV. 1501, 1512 (2002).

[314] Zacharias, *supra* note 311, at 999.

[315] *See generally* Leslie C. Levin, *Bad Apples, Bad Lawyers and Bad Decisionmaking: Lessons from Psychology and from Lawyers in the Dock*, 22 GEO. J. LEGAL ETHICS 1549, 1552 & n.30 (2009) (book review) ("As a group, lawyers tend to be more aggressive, competitive and achievement-oriented than the average individual." (citing SUSAN SWAM DELCOFF, LAWYER KNOW THYSELF: A PSYCHOLOGICAL ANALYSIS OF PERSONALITY STRENGTHS AND WEAKNESSES, 26-28 (2004)).

[316] *See generally* Jona Goldschmidt, *How Do Lawyer Disciplinary Agencies Enforce Rules Against Litigation Misconduct? Or Do They? Result of a Case Study and National Survey of Disciplinary Counsel*, 27 SUFFOLK J. TRIAL & APP. ADVOC. 1 (describing the studies and surveys of lawyer discipline by state disciplinary agencies nationwide and concluding that there is neither empirical evidence or appetite for enforcement of such agencies to discipline lawyers for their litigation conduct).

[317] *See, e.g.*, Monroe H. Freedman, *Professional Responsibility of the Criminal Defense Lawyer: The Three Hardest Questions*, 64 MICH. L. REV. 1469, 1474–82 (1966) (arguing that a defense attorney should not hesitate to zealously cross-examine truthful witnesses, should be permitted to call to the stand a client who intends to testify falsely, and ethically may give legal advice to a client even when doing so will tempt the client to commit perjury); Murray L. Schwartz, *On Making the True Look False and the False Look True*, 41 SW. L.J. 1135, 1143–44 (1988); Charles W. Wolfram, *Client Perjury*, 50 S. CAL. L. REV. 809, 854–55 (1977).

a criminal defense exemption exist in the Model Rules.[318] The provisions concerning the duty to investigate,[319] confining one's advocacy to the bounds of the law,[320] the prohibition against false evidence,[321] declining representation that would result in a violation of law or other fraud,[322] exercising independent judgment that includes moral factors,[323] and acting in good faith[324] make no distinction between criminal and civil cases.[325] Making the prosecution overcome false evidence is not just another legitimate way of putting the prosecution to its proof. Though the criminal defense lawyer is permitted to vigorously defend against criminal charges, even on behalf of an admittedly guilty client, relying on false evidence or arguing that true evidence is false, as Guy Reffitt's lawyer did when he deployed the deepfake defense, does no such thing.[326] It does not simply put the prosecution to its proof. "(I)t is one thing to attack a weak government case by pointing out its weakness. It is another to attack a strong government case by confusing the jury with falsehoods."[327] Nor does the right to effective counsel command otherwise.[328] Lawyers' duties to their clients in civil and criminal matters must be satisfied in conjunction with rather than in opposition to other obligations imposed by the law.

Therefore, the professional conduct rules, the substantive law, and the rules of procedure would not invariably or expressly subject lawyers to discipline or sanction for the use of the deepfake defense. Even when representing clients or otherwise conducting themselves as professionals, lawyers have some latitude to be untruthful. The outcomes of various attorney misconduct cases demonstrate a lack of consensus on whether and when attorneys may ethically engage in deception. This appears true even in cases where the ends seem to justify the means—when the lesser of two evils seems to justify the use of attorney deception. But case-by-case determinations on the permissibility of

---

[318] *Compare* MODEL RULES OF PRO. CONDUCT r. 3.1 (AM. BAR ASS'N 2021) (making a distinction between criminal and civil cases), *with id.* r. 3.3 (making no distinction).

[319] *See id.* r. 1.1.

[320] *Id.* at pmbl. 5 & 9.

[321] *Id.* 3.3(a)(3) The commentary notes that there is a debate about whether a criminal defense lawyer is under a different obligation, but it does not endorse the view, instead stating that the general rule—that an advocate must disclose the existence of perjury with respect to a material fact, even that of the client—"appl[ies] to all lawyers, including defense counsel in criminal cases." *Id.* cmt. 7.

[322] *Id.* r. 1.16.

[323] *Id.* at r. 2.1.

[324] *See* MODEL RULES OF PRO. CONDUCT r. 3.4(e).

[325] Model Rule 3.1 does make a distinction, permitting a criminal defense lawyer to plead not guilty despite knowledge that the client is guilty and require the state to prove guilt beyond a reasonable doubt. *Id.* r. 3.1.

[326] Verkouteren, *supra* note 6.

[327] Harry I. Subin, *The Criminal Lawyer's "Different Mission": Reflections on the "Right" to Present a False Case*, 1 GEO. J. LEGAL ETHICS 125, 148 (1987).

[328] *See* J. Alexander Tanford, *The Ethics of Evidence*, 25 AM. J. TRIAL ADVOC. 487, 509 (2002).

attorney deception threaten to slide along a slippery slope, leading to public mistrust of attorneys, attorney confusion regarding ethical expectations, and, ultimately, either unenforceable or under-enforced ethical standards. Indeed, if each attorney is free to carve out exceptions to deception prohibitions unilaterally, then ethics rules cease to function as rules; they become mere guidelines. To safeguard the integrity of the legal profession and protect attorneys, there must be sound, clear, enforceable legal and ethical standards that effectively govern the conduct of attorneys who use deception like the deepfake defense.

IV. PROPOSED SOLUTIONS

The deepfake defense challenges the integrity of legal proceedings and undermines the court's truth-finding process. The effect of deepfakes on the courts and trial processes has received some attention from scholars, law students, journalists, and practitioners.[329] Although the existing legal commentary acknowledges the threat of deepfakes in court proceedings in general, the exclusive focus of the conversation has been on the presentation and admission of evidence[330] rather than lawyers' efforts to exploit the idea of deepfakes to mislead and distract the jury. Thus, the solutions offered in the scholarship and articles are general and offer no specific guidance on how to curb the use of the deepfake defense. As this part explains, the problem of the deepfake defense requires a multi-dimensional response. These seeds of the solutions are in the existing rules of procedure, substantive law, and ethics.

---

[329] *See* Delfino, *Deepfakes on Trial*, *supra* note 13, at 339–40 (summarizing the scholarship).

[330] *See, e.g.*, *id.* at 339–48 (arguing that countering juror skepticism and doubt about the authenticity of audiovisual images in the era of fake news and deepfakes requires the reallocation of the fact-finding authority from the jury to the trial judge to determine the authenticity of audiovisual evidence, and proposing Federal Rule of Evidence 901 should be amended to add a new subdivision (c) to expand the gatekeeper functions of the court by assigning the responsibility to decide authenticity issues solely to the judge); *see also* Jonathan Mraunac, *The Future of Authenticating Audio and Video Evidence*, LAW360 (July 26, 2018), https://www.law360.com/articles/1067033/the-future-of-authenticating-audio-and-video-evidence [https://perma.cc/2DAE-7Y24]; Claypoole, *supra* note 306; Ashley Dean, *Deepfakes, Pose Detection, and the Death of "Seeing Is Believing*," L. TECH. TODAY (Aug. 6, 2020), https://www.lawtechnologytoday.org/2020/08/deepfakes-pose-detection-and-the-death-of-seeing-is-believing/ [https://perma.cc/GYB9-JAGM]; Lehman, Edson & Smith, *supra* note 55; Marie-Helen Maras & Alex Alexandrou, *Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and in the Wake of Deepfake Videos*, 23 INT'L J. EVIDENCE & PROOF 255, 255–62 (2019); Jason Tashea, *As Deepfakes Make It Harder to Discern Truth, Lawyers Can Be Gatekeepers*, Am. Bar Ass'n J. (Feb. 26, 2019), http://www.abajournal.com/lawscribbler/article/as-deepfakes-make-it-harder-to-discern-truth-lawyers-can-be-gatekeepers [https://perma.cc/38VS-DCRC].

A. *Solutions Grounded in the Rules of Procedure*

Even if the court decides all issues of authenticity and admissibility of audiovisual evidence outside the presence of the jury and, after that, instructs the jury to accept that the evidence is genuine—that is to assume that the evidence is not a deepfake—some lawyers may still use the deepfake defense to exploit the existence of deepfakes in their arguments to the court and jury. Sometimes this effort may be met by employing FRCP Rule 11 sanctions. However, as discussed in Part III, Rule 11 sanctions are authorized only where the lawyer's conduct relates to a signed writing. If the counsel's conduct does not manifest itself in writing, then FRCP Rule 11 would not apply.

Theoretically, a trial court could deter and punish counsel for using the deepfake defense using the court's inherent authority to control and manage proceedings before it. However, as discussed in Part III, courts have shown a hesitancy to rely exclusively on that authority to sanction lawyers' misconduct.[331] Therefore, providing the court with additional tools under Rule 11 would present one solution. Specifically, the definition of sanctionable conduct under Rule 11 should include lawyer conduct that manifests it.

California law offers a model for a revision to Rule 11. California Code of Civil Procedure section 128.7 mirrors Rule 11.[332] And like Rule 11, California's

---

[331] *See supra* Part II.B.1.a.

[332] California Code of Civil Procedure Section 128.7 provides, in pertinent part:

> (b) By presenting to the court, whether by signing, filing, submitting, or later advocating, a pleading, petition, written notice of motion, or other similar paper, an attorney or unrepresented party is certifying that to the best of the person's knowledge, information, and belief, formed after an inquiry reasonable under the circumstances, all of the following conditions are met: (1) It is not being presented primarily for an improper purpose, such as to harass or to cause unnecessary delay or needless increase in the cost of litigation. (2) The claims, defenses, and other legal contentions therein are warranted by existing law or by a nonfrivolous argument for the extension, modification, or reversal of existing law or the establishment of new law. (3) The allegations and other factual contentions have evidentiary support or, if specifically so identified, are likely to have evidentiary support after a reasonable opportunity for further investigation or discovery. (4) The denials of factual contentions are warranted on the evidence or, if specifically so identified, are reasonably based on a lack of information or belief. (c) If, after notice and a reasonable opportunity to respond, the court determines that subdivision (b) has been violated, the court may, subject to the conditions stated below, impose an appropriate sanction upon the attorneys, law firms, or parties that have violated subdivision (b) or are responsible for the violation. In determining what sanctions, if any, should be ordered, the court shall consider whether a party seeking sanctions has exercised due diligence.

CAL. CIV. PROC. CODE § 128.7 (West 2023).

section 128.7 is limited in scope, providing authority to sanction lawyers whose misconduct can be connected to a signed "pleading, petition, written notice of motion, or other similar paper . . . ."[333] However, unlike Rule 11, section 128.7 does not stand alone. In 2014, the California Legislature enacted section 128.5, which authorizes the court to sanction counsel (and the parties) for a broader range of conduct than section 128.7, including "actions or tactics, made in bad faith, that are frivolous or solely intended to cause unnecessary delay."[334] Thus, section 128.5 is not limited to conduct memorialized in writing—it applied to "bad-faith actions or tactics."[335] The statute further defines "frivolous" as conduct "completely without merit or for the sole purpose of harassing an opposing party,"[336] and Section 128.5 is not limited to written conduct.[337] Section 128.5 was enacted because standing alone, section 128.7 had not proven effective in halting lawyers' misconduct, including oral misconduct.[338] Rule 11

---

[333] *See id.* § 128.7(b).

[334] *Id.* § 128.5.

[335] Section 128.5. provides in pertinent part:

> (a) A trial court may order a party, the party's attorney, or both, to pay the reasonable expenses, including attorney's fees, incurred by another party as a result of actions or tactics, made in bad faith, that are frivolous or solely intended to cause unnecessary delay. This section also applies to judicial arbitration proceedings under Chapter 2.5 (commencing with Section 1141.10) of Title 3 of Part 3.
> (b) For purposes of this section: (1) "Actions or tactics" include, but are not limited to, the making or opposing of motions or the filing and service of a complaint, cross-complaint, answer, or other responsive pleading. The mere filing of a complaint without service thereof on an opposing party does not constitute "actions or tactics" for purposes of this section. (2) "Frivolous" means totally and completely without merit or for the sole purpose of harassing an opposing party.(c) Expenses pursuant to this section shall not be imposed except on notice contained in a party's moving or responding papers or, on the court's own motion, after notice and opportunity to be heard. An order imposing expenses shall be in writing and shall recite in detail the action or tactic or circumstances justifying the order. (d) In addition to any award pursuant to this section for an action or tactic described in subdivision (a), the court may assess punitive damages against the plaintiff on a determination by the court that the plaintiff's action was an action maintained by a person convicted of a felony against the person's victim, or the victim's heirs, relatives, estate, or personal representative, for injuries arising from the acts for which the person was convicted of a felony, and that the plaintiff is guilty of fraud, oppression, or malice in maintaining the action.

Id. § 128.5(a)–(e).

[336] *Id.* § 128.5(b)(2).

[337] *Id.* § 128.5(b)(1).

[338] *See California Revives Old Attorney Sanctions Statute*, LEGIS. INTENT SERV., INC., http://www.legintent.com/california-revives-old-model-attorney-sanctions/ [https://perma.cc/

suffers from the same inadequacy as section 128.7. Taking the language of section 128.5 as a guide, Rule 11 should be revised to expressly expand its reach to actions and tactics that manifest in oral argument.[339]

B. *Ethical Rule Remedies*

In addition to expanding the reach of Rule 11 to apply to oral argument, the deepfake defense calls for a revision to the rules of professional conduct governing lawyers. A revision to ABA Model Rule 3.3 is an excellent place to start. Rule 3.3 should be revised to address deepfake evidence and the deepfake defense by adding a new subdivision (b).[340] The new subdivision (b) should provide:

> Rule 3.3(b): Deepfake Evidence in Court:
> 1. A lawyer shall not knowingly or recklessly:
> (a) Offer evidence that the lawyer knows or should have known after reasonable due diligence to be a deepfake; or
> (b) Impugn the authenticity of evidence without reason to believe that the evidence is a deepfake or argue or imply that evidence is false unless the lawyer has a reasonable belief that the evidence is false.
> 2. If a lawyer, the lawyer's client, or a witness called by the lawyer, has offered deepfake evidence and the lawyer comes to know of its falsity, the lawyer shall take reasonable remedial measures, including, if necessary, disclosure to the tribunal. A lawyer may refuse to offer evidence, other than

---

C2G2-5LXJ]. The California State Assembly Member, Ken Cooley reflected on the need for section 128.5:

> Unfortunately, bad-faith disobedience and tactics by either side are needlessly employed in litigation. It can result in clogging our courts and wasting precious judicial resources. Courts routinely give litigants the benefit of the doubt, but they have lost an important tool used to ensure bad faith actions that can materially harm the other party or the fairness of a trial are discouraged. Under existing law a court can compel obedience with a court order, but financially the most a court can do if a party violates one is find them in contempt with penalty of up to $1500. Moreover, if a case ends in a mistrial or in the release of protected documents it is difficult to undo the waste of judicial resources or harm done to the litigant who was not at fault.

Id.

[339] Courts have imposed Section 128.5 sanctions based on oral misrepresentations made during the court proceedings. *See, e.g.*, *In re* Marriage of Sahafzadeh-Taeb & Taeb, 39 Cal. App. 5th 124, 128, 132 (2019) (upholding a sanctions order under section 128.5 based on the attorney's oral misrepresentation to the court that she would be ready to proceed on the scheduled trial date).

[340] The current subdivisions, (b) through (d) should be retained but redesignated as (c) through (e).

> the testimony of a defendant in a criminal matter, that the lawyer reasonably believes is a deepfake.[341]

To account for the seriousness of introducing deepfake evidence into court, this proposed rule holds attorneys accountable for their knowing and reckless use of the deepfake defense. This rule forces an attorney to go beyond just knowing whether their evidence is deepfake and to take the additional step of conducting a reasonable analysis of their evidence to ensure it is legitimate.

Moreover, subsection (b)(2) references an attorney's ethical duty not to attack evidence without actual cause to believe the evidence is a deepfake. Including this language will force attorneys to challenge evidence only in those situations where they reasonably believe they must do so. To that end, the new Rule 3.3, subdivision (b)(1)(b), requires that a lawyer refuse to *argue* or suggest that evidence is false unless the lawyer has a reasonable belief that the evidence is false. This express and unequivocal ethical prohibition would discourage lawyers from arguing that evidence is fake unless they have a reasonable basis for doing so.

C. *Substantive Legal Responses*

Because, as previously discussed, both the courts and bar regulators have not consistently punished lawyer misconduct, solutions to curb the deepfake defense grounded exclusively in the court's inherent discretionary power, application of the rules of court, or ethics may not prove a potent response. Thus, a statutory remedy should also be considered. Amending 28 U.S.C. § 1927[342] offers another answer to counter the deepfake defense because of its broad application to criminal and civil proceedings. Unlike Rule 11, 28 U.S.C. § 1927 does not depend on a signed writing.[343] However, its contours are not clearly defined.[344] Because the appellate courts do not agree on the type of conduct that constitutes an "unreasonable and vexatious multiplication of the proceedings," Section 1927 is not uniformly applied.[345] Thus, Section 1927 should be amended to provide a clear and consistent definition, clarifying that offering

---

[341] The new subdivision (b) would also require the creation of a Comment to define deepfake in this context as: The term "deepfake" means an audiovisual record created or altered in a manner that the record would falsely appear to a reasonable observer to be an authentic record of the actual speech, conduct, image, or likeness of an individual.

[342] Under Section 1927 any attorney "[W]ho so multiplies the proceedings in any case unreasonably and vexatiously may be required by the court to satisfy personally the excess of costs, expenses, and attorneys' fees reasonably incurred because of such conduct." 28 U.S.C. § 1927.

[343] *See id.*

[344] *See* Josselyn, *supra* note 192, at 481 (describing the case law interpreting section 1927 and lamenting that the courts do not apply the standard for assessing sanctionable conduct under the statute in a clear and consistent manner).

[345] *See supra* Part III.

deepfake evidence or deploying the deepfake defense constitutes unreasonable and vexatious litigation conduct that Section 1927 was designed to prohibit.

Critics of these proposals may be concerned that prohibiting lawyers from using the deepfake defense would undermine counsel's ability to advocate on behalf of clients zealously. Not so. The ideas offered here do not undermine a lawyer's ability to employ deception to represent their clients in all instances.[346] Others might suggest that this regulation of lawyers offends constitutional norms and values. Although lawyers do not relinquish their First Amendment rights when they obtain a law license, lawyers can nonetheless be subject to speech restrictions, such as evidentiary and ethics rules that restrict what they can say in court.[347] Lawyers' speech in court and in the context of litigation can be regulated to protect the integrity of judicial proceedings.[348] There is no separate category of professional speech that is, by its nature, subject to regulation. But states can regulate lawyers' conduct when necessary to ensure the proper functioning of the judicial process, even if the regulation directly or incidentally interferes with speech.[349]

Others may question whether it is necessary to single out deepfake evidence and use the deepfake defense for special sanctions. But deepfakes are truly different from other evidence arguments and defenses that lawyers deploy. By design, deepfakes trick the viewer; they raise existential questions about "reality" on a profound and metaphysical level.[350] Deepfakes, fake news, and conspiracy theories have invaded the cultural ethos and political pathos.[351] Moreover, the leading digital forensic experts worry that the fight to detect deepfakes is a losing battle—that the deepfaker's technology is outstripping the ability of those trying to detect the deepfakes.[352] Thus, deepfakes and using the deepfake defense present an exceptional and unprecedented challenge to our legal system that warrants unique treatment under the law and ethical standards. Deepfakes have emerged and thrived in an era where all truth is under siege. Deepfakes exploit the divide between those who believe that facts exist and those who only believe in a reality that confirms their own biases. When jurors are presented with the idea that evidence may be a deepfake, there is a genuine

---

[346] *See* Jefferson, *supra* note 123, at 125–27 (describing the range of acceptable lying for lawyers).

[347] *E.g.*, MODEL CODE OF PRO. CONDUCT r. 3.3, 3.4(b), 3.5 (AM. BAR ASS'N 2021).

[348] *See* Gentile v. State Bar of Nev., 501 U.S. 1030, 1071 (1991) ("It is unquestionable that in the courtroom itself, during a judicial proceeding, whatever right to 'free speech' an attorney has is extremely circumscribed.").

[349] Nat'l Inst. of Fam. & Life Advocs. v. Becerra, 138 S. Ct. 2361, 2365–66 (2018) ("[T]his Court has never recognized 'professional speech' as a separate category of speech subject to different rules.").

[350] *See supra* Part II.A.

[351] *See, e.g.*, Chesney & Citron, *supra* note 19, at 1755–58.

[352] *See* Brown, *supra* note 54, at 25 (discussing the belief among the community of computer science and digital forensics experts that detection methods cannot keep pace with the innovations aimed at evading detection).

risk that a jury will harbor doubts about what they see and hear either based on a "seeing is believing mindset" or the opposite—a view that everything is fake and that those biases are so strong that they will taint their deliberative process.[353] Since lawyers play a unique role in the justice system and may be viewed by juries as possessing insider information about their cases,[354] the deployment of the deepfake defense thus poses a particular danger because it foments the cognitive dissonance and confirmation bias that allows deepfakes to flourish.[355] Thus, the deepfake defense threatens our democratic institutions, including our justice system.

The solutions offered here may not be the last word on how to best confront the challenges the deepfake defense presents. But given that we live in the era of the daily bombardment by fake news, conspiracy theories, technological trickery, and deepfakes which triggers juries to question and distrust even genuine evidence, these proposals can begin the conversation.

V. CONCLUSION

The American legal adversarial system's effective functioning depends on lawyers advocating on behalf of their clients with zeal but within the bounds of the law and ethical rules. It also depends on the trier of fact—either the judge or the jury—finding the truth based on the evidence presented. As deepfake audio-digital images proliferate in society, they have inevitably invaded legal proceedings. This invasion presents new challenges for juries who will have a harder time detecting truthful evidence from fake evidence as they confront lawyers' use of the deepfake defense to falsely claim that the evidence presented at trial is fake.

Deepfakes will require lawyers and courts to take additional measures and employ additional resources to determine the authenticity of images before presenting them as evidence. And they will also require the court to deter lawyers from exploiting the deepfake defense when the lawyer knows the

---

353 *See* Chesney & Citron, *supra* note 19, at 1777–78.

354 Trial practice literature on jurors' perception of trial lawyers underscores that juries place special faith in trial lawyers as interpreters of the law and evidence. *See* Robert L. Hollingshead & John C. Maloney, Jr., *The Opening Statement*, N.J. LAW., Dec. 1998, at 16, 17 ("[W]hile recognizing that the attorneys have different views of the evidence, [jurors] trust that the attorneys will tell them the truth regarding what the evidence will demonstrate in support of those respective views. They especially will trust the attorney whose client they believe is 'in the right.'"); Latour "LT" Lafferty, *Leadership in Trial Advocacy: Credibility Is a Cornerstone of Effective Trial Advocacy*, 28 AM. J. TRIAL ADVOC. 517, 527 (2005) (explaining that jurors trust lawyers that have integrity and credibility, and that jurors ultimately vote for the lawyer that they "believe in"); Harry J. Plotkin, Feature, *Building Trust Among the Jury Creating Positive Impressions of Witnesses and Attorneys . . .* , 47 ORANGE CNTY. LAW. 28, 28 (2005) (positing that juries do not make their decisions solely based on evidence, but rather that they consider their trust in attorneys when making their decisions).

355 *See supra* Part II.C.

evidence is genuine to confuse and mislead the jury. As lawyers, courts, and jurors traverse this new technological landscape of deepfakes, confronting and stopping the deepfake defense will require a multi-direction response from the rules of procedure, substantive law, and ethics. Deploying the new tools and mechanisms will protect the jury's truth-seeking function upon which our justice system depends.