Deepest Fakes

Mihailis E. Diamantis,[*] Sean Sullivan[**] & Eli Alshanetsky[***]

Abstract

Deepfakes are visual and audio media that use artificial intelligence to portray people saying things they never said, doing things they never did, and experiencing events that never happened. Because deepfakes can be both persuasive and pervasive, many commentators fear that humanity will soon take another step into the post-truth abyss.

This Article evaluates the threat deepfakes pose to truth by anticipating how they will impact the area of law most directly concerned with truth: the law of evidence. Deepfakes present an obvious challenge to the administration of justice in modern courtrooms, where audiovisual evidence plays an important role. Solutions offered in past legal scholarship—like relying on experts to identify deepfakes or criminalizing deepfake production—optimistically assume that deepfakes will always have a tell. To truly appreciate the threat deepfakes pose, the law must brace itself for the likely prospect of "deepest" fakes, which will be indistinguishable in every respect from authentic media.

Drawing on tools from philosophy, legal history, and technology studies, this Article demonstrates how evidence law can and likely will adapt to a world saturated with deepest fakes. This Article finds that deepest fakes present no different challenge for modern courts than oral testimony, paintings, photographs, and other easily falsifiable evidence presented for their early twentieth-century counterparts. The safeguard then, as now, is a nuanced adversarial process that refuses to take evidence at face value and probes each submission with contextual indicators of reliability. What emerges is an empowering picture in which human judgment, rather than blind trust in media, is the ultimate arbiter of truth.

Table of Contents

[*] Ben V. Willie Professor in Excellence, University of Iowa College of Law.

[**] Jon and Sarah Fister Chair in Law & Economics, University of Iowa College of Law.

[***] Assistant Professor in Philosophy, Temple University.

The Authors are grateful to Hunter Thompson for research support.

February 2026 Vol. 94 No. 1

Don't believe everything you read on the internet.

–Abraham Lincoln (circa 1864)

Introduction

Deepfakes splashed into the 2024 election like never before.[1] Residents in New Hampshire received robocalls in which President Biden urged them not to turn out for the election.[2] Digital images showed Taylor Swift, dressed as Uncle Sam, endorsing Donald Trump.[3] A photo on X showed a younger Trump groping a minor with sex offender

[1] *See generally The AI-Generated Hell of the 2024 Election*, Verge (Nov. 3, 2024, at 19:36 ET), https://www.theverge.com/policy/24098798/2024-election-ai-generated-disinformation [https://perma.cc/LC8S-ECRG] (collecting articles discussing artificial intelligence and the 2024 election cycle). Perhaps surprisingly, we have it relatively easy here in the United States. In the United Kingdom, there are claims that politicians themselves are deepfakes. Mia Sato, *The UK Politician Accused of Being AI Is Actually a Real Person*, Verge (July 9, 2024, at 16:45 ET), https://www.theverge.com/2024/7/9/24195005/reform-uk-candidate-election-ai-bot-mark-matlock [https://perma.cc/XCZ4-Q8E5].

[2] Lauren Feiner, *Telecom Will Pay $1 Million over Deepfake Joe Biden Robocall*, Verge (Aug. 21, 2024, at 15:27 ET), https://www.theverge.com/2024/8/21/24225435/lingo-telecom-biden-deepfake-robocall-fcc-fine [https://perma.cc/59B5-FC8Q].

[3] Shannon Bond, *How AI-Generated Memes Are Changing the 2024 Election*, NPR (Aug. 30, 2024, at 05:00 ET), https://www.npr.org/2024/08/30/nx-s1-5087913/donald-trump-artificial-intelligence-memes-deepfakes-taylor-swift [https://perma.cc/FF78-K48R].

Jeffrey Epstein.[4] Biden wearing military fatigues.[5] Trump embraced by Black voters.[6]

None of it was real. The consultant behind the robocalls faced criminal charges and a hefty fine.[7] Swift later endorsed Harris.[8] The other images were all debunked.[9] But even in their absence, deepfakes still drove news, as when Trump falsely claimed that photos of Vice President Harris's huge rally crowds were fake.[10]

Deepfakes are audiovisual media that use deep learning to seamlessly stitch together faces, voices, and other elements into highly realistic representations.[11] They are "fake" at two levels: their content (they portray events that never happened) and their presentation (they deceptively appear to be traditional recordings captured by mechanical devices like cameras). Deepfakes first gained attention on Reddit in 2017 as pornographic videos that swapped celebrities in for the true actors.[12] Shortly after, Lyrebird debuted, giving people a way "to recreate anyone's voice and get it to say almost anything."[13] A visual media program called FakeApp launched the same year with the explicit goal of making deepfake technology "available to people without a technical

[4] Aleksandra Wrona, *Does Pic Show Trump and Epstein with Minor Girl?*, Snopes (Jan. 9, 2024), https://www.snopes.com/fact-check/epstein-trump-young-girl-photo/ [https://perma.cc/9AQJ-MS38].

[5] Bill McCarthy, *Image of Biden Planning Military Action in Fatigues Is Fake*, AFP Fact Check (Apr. 29, 2024, at 11:13 ET), https://factcheck.afp.com/doc.afp.com.34H74GF [https://perma.cc/8SYM-EY4X].

[6] Marianna Spring, *Trump Supporters Target Black Voters with Faked AI Images*, BBC (Mar. 4, 2024), https://www.bbc.com/news/world-us-canada-68440150 [https://perma.cc/5HZP-9LTS].

[7] Shannon Bond, *A Political Consultant Faces Charges and Fines for Biden Deepfake Robocalls*, NPR (May 23, 2024, at 14:58 ET), https://www.npr.org/2024/05/23/nx-s1-4977582/fcc-ai-deepfake-robocall-biden-new-hampshire-political-operative [https://perma.cc/5X4Z-YS2S]; Lingo Telecom, LLC, 39 FCC Rcd. 9304, 9304 (2024).

[8] Chloe Veltman, *Taylor Swift Has Endorsed Kamala Harris for President—Will It Matter?*, NPR (Sep. 11, 2024, at 12:32 ET), https://www.npr.org/2024/09/11/nx-s1-5108695/taylor-swift-endorsement-kamala-harris [https://perma.cc/X4JP-E5GZ].

[9] *See* sources cited *supra* notes 4–6.

[10] Jude Joffe-Block, *Why False Claims that a Picture of a Kamala Harris Rally Was AI-Generated Matter*, NPR (Aug. 14, 2024, at 17:26 ET), https://www.npr.org/2024/08/14/nx-s1-5072687/trump-harris-walz-election-rally-ai-fakes [https://perma.cc/KX5D-4V6L].

[11] Douglas Harris, Article, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 Duke L. & Tech. Rev. 99, 99–100 (2019); Rebecca A. Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 Fordham L. Rev. 887, 889, 892–93 (2019).

[12] 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, 299 (2023).

[13] Morning Edition: *New Software Can Mimic Anyone's Voice* (NPR, May 5, 2017), http://www.npr.org/2017/05/05/527013820/ [https://perma.cc/2SBS-4GTN].

background or programming experience."[14] Since then, the beneficial and nefarious uses of this technology have been limited only by users' imaginations.

Deepfakes leave many people feeling rattled. Of course, there are obvious harmful applications for deepfakes, like stealing people's identities, creating unauthorized pornography, and engaging in fraudulent transactions. But some commentators foretell a broader, more structural threat. According to them, deepfakes could sow social discord: "A well-timed and thoughtfully scripted deep fake . . . could tip an election, spark violence in a city primed for civil unrest, . . . or exacerbate political divisions in a society."[15] More worrying, deepfakes could "impact . . . the very fabric of our democracy":[16] "[T]he informational anarchy and paranoia that [deepfakes cause] . . . might . . . challenge . . . individual decision making or collective self-rule . . . ."[17] Finally and most extreme, deepfakes could wage "war on reality"[18] because "[t]hey raise existential questions about reality on a profound and metaphysical level."[19] For some commentators, "war" is not a metaphor; they suggest that certain deepfakes could warrant "a military response."[20]

Our starting observation is this: Saying that "deepfakes raise existential questions about reality"[21] or that "people [can] no longer believe *anything* is real [if they cannot trust digital media]"[22] conflates reality with audiovisual representations of it. Of course, we can only know what is real if we have evidence of it. In a world in which the average adult spends most of their waking life looking at a screen,[23] it may

[14] *See* Samantha Cole, *We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*, VICE (Jan. 24, 2018, at 14:13 ET), https://www.vice.com/en/article/bjye8a/reddit-fake-porn-app-daisy-ridley [https://perma.cc/EY73-34A2].

[15] Robert Chesney & Danielle K. Citron, *Disinformation on Steroids: The Threat of Deep Fakes*, COUNCIL ON FOREIGN RELS. (Oct. 16, 2018), https://www.cfr.org/report/deep-fake-disinformation-steroids [https://perma.cc/U3RZ-SY6A].

[16] Robert Chesney & Danielle Citron, *Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?*, LAWFARE (Feb. 21, 2018, at 10:00 ET), https://www.lawfaremedia.org/article/deepfakes-looming-crisis-national-security-democracy-and-privacy [https://perma.cc/W5PD-RLUD].

[17] Marc Jonathan Blitz, *Lies, Line Drawing, and (Deep) Fake News*, 71 OKLA. L. REV. 59, 110 (2018).

[18] Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J.L. & TECH. 1, 8 (2020).

[19] Delfino, *supra* note 12, at 345.

[20] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1809 (2019).

[21] Rebecca A. Delfino, *The Deepfake Defense—Exploring the Limits of the Law and Ethical Norms in Protecting Legal Proceedings from Lying Lawyers*, 84 OHIO ST. L.J. 1067, 1081 (2024).

[22] Agnieszka McPeak, *The Threat of Deepfakes in Litigation: Raising the Authentication Bar to Combat Falsehood*, 23 VAND. J. ENT. & TECH. L. 433, 439 (2021).

[23] *See* Brian Stelter, *8 Hours a Day Spent on Screens, Study Finds*, N.Y. TIMES (Mar. 26, 2009), https://www.nytimes.com/2009/03/27/business/media/27adco.html [https://perma.

be easy to mix up reality and the projection of it. We agree that deepfakes present "a fundamental challenge to the [existing] information environment."[24] But it is important to recognize that digital media are only one source of evidence, and they have not even been around for that long. Although it may be hard to remember a time before everyone had Snapchat in their pocket, digital cameras were not commercially available until 1990,[25] having first been invented fifteen years prior.[26]

In this Article, we—a law and tech scholar, an evidence scholar, and an epistemologist—defend both a thesis and an antithesis. The thesis is that deepfakes pose even greater problems than commentators envision. The antithesis is that existing social, legal, and metaphysical institutions for responding to creative forms of lying are more powerful than commentators realize. Epistemology and the law of evidence help us to recall the various ways that humans guard against the possibility of deception and the ways that reality reveals itself to skeptical audiences, even when once-reliable sources of information cease to be trustworthy.[27] Indeed, it is possible that deepfakes will ultimately empower people for the very reasons that most commentators see them as a threat. By undermining the reliability of digital media, deepfakes may emphasize and reaffirm the ultimate authority of human judgment.

To make our case, we turn to the law's most truth-focused institution: the courtroom, which has been called "a microcosm of society in general."[28] By assessing the impact deepfakes will have on courts' truth-finding mission, we glean lessons for society more generally. Deepfakes are already appearing as evidence in trial,[29] and scholars

cc/39J7-E2Q7]; Jacqueline Howard, *Americans Devote More Than 10 Hours a Day to Screen Time, and Growing*, CNN (July 29, 2016, at 16:22 ET), https://www.cnn.com/2016/06/30/health/americans-screen-time-nielsen [https://perma.cc/PN26-P7BM]; *Average U.S. Adult Will Spend Equivalent of 44 Years of Their Life Staring at Screens: Poll*, People (June 3, 2020, at 16:46 ET), https://people.com/human-interest/average-us-adult-screens-study/ [https://perma.cc/CJV2-S9Z8].

[24] Mark Corcoran & Matt Henry, *The Tom Cruise Deepfake That Set Off 'Terror' in the Heart of Washington DC*, ABC News Austl. (June 27, 2021, at 20:16 ET), https://www.abc.net.au/news/2021-06-24/tom-cruise-deepfake-chris-ume-security-washington-dc/100234772 [https://perma.cc/WR6Q-F8CC] (quoting Matt Ferraro, a former agent of the Central Intelligence Agency).

[25] Lauren Cabral, *The History of Digital Cameras*, Back Then Hist. (July 27, 2023), https://www.backthenhistory.com/articles/the-history-of-digital-cameras [https://perma.cc/KU2R-ZRLR].

[26] Joanna Goodrich, *The First Digital Camera Was the Size of a Toaster*, IEEE Spectrum: Institute (Apr. 6, 2022), https://spectrum.ieee.org/first-digital-camera-history [https://perma.cc/D5BU-JNLT].

[27] As the Federal Rules of Evidence say, their goal is to help courts with "the end of ascertaining the truth" from available information. Fed. R. Evid. 102.

[28] Riana Pfefferkorn, *"Deepfakes" in the Courtroom*, 29 B.U. Pub. Int. L.J. 245, 257 (2020).

[29] *See, e.g.*, Matt Reynolds, *Courts and Lawyers Struggle with Growing Prevalence of Deepfakes*, ABA J. (June 9, 2020, at 09:29 CT), https://www.abajournal.com/web/article/courts-and-lawyers-struggle-with-growing-prevalence-of-deepfakes [https://perma.cc/VXM9-2L39].

are starting to propose various measures to preserve courts' integrity.[30] Existing recommendations assume that there will always be a sophisticated way to tell deepfakes and genuine media apart. But this leaves courts vulnerable to what this Article calls "deepest fakes"—deepfake media that technologists predict will be costless to make and will perfectly mimic genuine media.[31] The prospect of deepest fakes places the threat of deepfakes, both to courtrooms and to broader institutions, in the starkest possible terms. Any approach that solves the problem of deepest fakes in courtrooms could likewise solve the problem of deepfakes in the wild.

For epistemologists, skeptical challenges to our grip on reality are as old as the discipline itself.[32] Epistemology offers conceptual tools for evaluating sources of evidence and overcoming skepticism. Many of these tools have analogues in evidence law.[33] The constant possibility that any photo, video, or audio recording could be faked recalls courts' historic struggles to separate truthful from dishonest testimony. We anticipate that courts will respond as they have in the past.[34] Minimal procedural adjustments may help, but the key safeguard against both deepfakes and deepest fakes is a robust adversarial process that provides jurors not only with digital media evidence but also with the contextual factors that bear on its veracity. The most important change will not occur within courts, but within jurors as they become more astute judges of media evidence. The implications for courts and broader society are empowering.[35] As we collectively learn that we can no longer reflexively trust digital media, our own evolving epistemic practices will grow to take on a more central truth-finding role. In the long run, we foresee a future in which these developments will actually strengthen humans' relationship with reality.

I. The Real Problem of Deepfakes

Before building out our antithesis, this Part starts by presenting the threat that deepfakes pose. That threat is greater than many realize. Legislative attempts to ban deepfakes in certain contexts[36] will not keep them from infiltrating the law's most truth-oriented institution:

[30] *See infra* Part I.

[31] *See* 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, 257 (2019) ("Put simply, when AI technology is used in the future, it may be impossible to determine that the video is fake.").

[32] *See infra* Part II.

[33] *See infra* Part III.

[34] *See infra* Part IV.

[35] *See infra* Conclusion.

[36] *See infra* Section I.A.

the courtroom.[37] Although evidence scholars have taken note,[38] their solutions rest on the misplaced optimism that deepfakes will always have some hidden tell that reveals them for what they are. Technologists are not so sanguine.[39] Any framework for responding to deepfakes must be resilient enough to withstand the eventuality of undetectable deepest fakes.

A. The Law of Deepfakes

The First Amendment presents a significant barrier to legislation that prohibits deepfakes. Deepfakes "are generally video or audio creations, and such creations have typically been considered a form of expression."[40] Deepfake legislation, by necessity, targets digital media that is fake. But as the Supreme Court opined fifty years ago, "Under the First Amendment there is no such thing as a false idea."[41] If the only possible uses of deepfake technology were social blights like blackmail and fraud, then the Constitution might not present a barrier to prohibiting them.[42] Part of the problem from a First Amendment perspective is that the same technology used to produce deepfakes can be put to honest and valuable uses, like portraying deceased actors in movie sequels[43] or enabling brain and throat cancer survivors to continue communicating in their own voice.[44]

Because of these First Amendment protections, the government can only regulate uses of deepfake technology that fall into a historically unprotected category of speech or expression.[45] There have been sporadic legislative efforts to control deepfakes in some particularly problematic contexts. The only federal legislation[46] on point is

[37] *See infra* Section I.B.

[38] *See infra* Section I.C.

[39] *See infra* Section I.D.

[40] Blitz, *supra* note 17, at 62.

[41] Gertz v. Robert Welch, Inc., 418 U.S. 323, 339 (1974).

[42] *See* United States v. Alvarez, 567 U.S. 709, 717–18 (2012) (plurality opinion).

[43] Peter Suciu, *Deepfake Star Wars Videos Portent Ways the Technology Could Be Employed for Good and Bad*, FORBES (Dec. 11, 2020, at 18:15 ET), https://www.forbes.com/sites/petersuciu/2020/12/11/deepfake-star-wars-videos-portent-ways-the-technology-could-be-employed-for-good-and-bad/ [https://perma.cc/P52B-WN6M].

[44] *See* Brooke Steinberg, *I Lost My Voice Because of a Tumor—but an AI Clone Gave It and My Confidence Back to Me*, N.Y. POST (May 13, 2024, at 13:57 ET), https://nypost.com/2024/05/13/lifestyle/i-lost-my-voice-because-of-a-tumor-an-ai-clone-gave-it-back-to-me/ [https://perma.cc/QJG2-CT75].

[45] *See Alvarez*, 567 U.S. at 717–18 (plurality opinion). In theory, a government could also draft regulation that satisfies strict scrutiny. *See id.* at 724–25.

[46] Other federal legislation has been proposed. For example, in 2018, Senator Ben Sasse introduced a bill that would have criminalized certain harmful deepfakes, Malicious Deep Fake Prohibition Act of 2018, S. 3805, 115th Cong. (2018), and in 2019, Representative Yvette Clarke introduced a bill that would have required deepfakes to carry watermarks, Defending Each and

the National Defense Authorization Act for Fiscal Year 2020,[47] which requires annual assessments of foreign efforts to weaponize deepfakes or to use them for election interference.[48] Some states have also passed limited criminal statutes,[49] like Virginia's law[50] against deepfake revenge porn or Texas's[51] and California's[52] laws against using deepfakes to influence elections. Whether these state statutes will survive the Supreme Court's First Amendment scrutiny remains uncertain.[53] One commentator has argued that statutes prohibiting deepfake pornography are likely unconstitutional.[54] Indeed, the Texas Court of Criminal Appeals held that its state law violated the First Amendment.[55] Additional challenges to Virginia's election statutes will surely arise in the litigious wake of the 2024 election cycle.[56]

Even if laws that specifically target deepfakes are on constitutionally shaky ground, several existing speech-neutral statutes already offer civil and criminal remedies for objectionable acts that could—but need not—involve deepfakes.[57] Laws in most states cover a range of harmful

Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R. 3230, 116th Cong. (2019). Both bills died in committee. *See All Information (Except Text) for S.3805—Malicious Deep Fake Prohibition Act of 2018*, Congress.gov, https://www.congress.gov/bill/115th-congress/senate-bill/3805/all-info [https://perma.cc/UV27-82KS] (last visited Nov. 17, 2025); *All Information (Except Text) for H.R.3230—DEEP FAKES Accountability Act*, Congress.gov, https://www.congress.gov/bill/116th-congress/house-bill/3230/all-info [https://perma.cc/4MEJ-77P7] (last visited Nov. 17, 2025).

[47] National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92, 133 Stat. 1198 (2019) (codified in scattered sections of U.S.C.).

[48] 50 U.S.C. § 3369(a).

[49] International efforts are underway too. For example, Germany has made it a crime to create a deepfake that violates personal rights. *See* Strafgesetzbuch [StGB] [Penal Code], Nov. 13, 1998, § 201a, last amended by Gesetz [G], Nov. 22, 2021, https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html [https://perma.cc/L98A-K2YZ].

[50] *See* Va. Code Ann. § 18.2-386.2 (2025); *see also* Cal. Civ. Code § 1708.86 (West 2025) (providing a private right of action to victims of deepfake pornography in California).

[51] Tex. Elec. Code Ann. § 255.004 (prohibiting use of a deceptive video with "intent to injure a candidate or influence the result of an election"), *invalidated by*, *Ex parte* Stafford, No. PD-0310-23, 2024 WL 4031614 (Tex. Crim. App. Sep. 4, 2024).

[52] *See* Cal. Elec. Code § 20010 (West 2025).

[53] *See* Bradley Waldstreicher, Note, *Deeply Fake, Deeply Disturbing, Deeply Constitutional: Why the First Amendment Likely Protects the Creation of Pornographic Deepfakes*, 42 Cardozo L. Rev. 729, 743 (2021).

[54] *See id.*

[55] *See Ex parte Stafford*, 2024 WL 4031614, at *8.

[56] *See 2023–2024 Litigation Report: The Most Litigated Election in History*, Democracy Dkt. (Dec. 11, 2024), https://www.democracydocket.com/analysis/2024-litigation-report/ [https://perma.cc/A9V4-MQW4] (referring to the 2024 election cycle as "the most litigated election cycle in history").

[57] *See* Project Veritas v. Schmidt, 72 F.4th 1043, 1062 n.15 (9th Cir. 2023), *vacated en banc*, 95 F.4th 1152 (9th Cir. 2024).

deepfake use cases, such as defamation, intentional infliction of emotional distress, impersonation, and cyberstalking, among others.[58] In a similar vein, Virginia's prior revenge porn statute applies to the unauthorized and malicious distribution of pornographic content "created *by any means whatsoever* that depicts another person."[59] Its broad language would seemingly include deepfake revenge porn, even without the later clarifying amendment: "'[A]nother person' includes a person whose image was used in creating, adapting, or modifying a videographic or still image with the intent to depict an actual person . . . ."[60] New Jersey is considering, but has yet to pass, its own deepfake statute.[61] As with Virginia, New Jersey's existing revenge porn statute is arguably broad enough to encompass deepfakes and neutral enough to satisfy the First Amendment.[62] It applies if someone "reproduces *in any manner*, the image of another person whose intimate parts are exposed . . . without that person's consent."[63]

B. Deepfakes in the Courtroom

Commentators are rightly skeptical of the effectiveness of legislation aimed at proscribing deepfakes. Statutory bans may provide some recourse for victims, but they will not stem the creation of deepfakes themselves.[64] Put another way, "We may safely assume that the ready availability of deepfake tools, and antisocial uses thereof, will continue irrespective of how the law may attempt to contain, regulate, and punish them."[65]

One emerging antisocial use of deepfake technology is to manipulate evidentiary records at trial: "[O]ur legal system is as vulnerable to content manipulation as any other area of civic life . . . ."[66] Indeed, although "[i]t is often illegal to make false statements where government needs honest answers to questions,"[67] courtrooms are already feeling the influence of deepfakes.

[58] *See, e.g.*, *id.* ("[V]ictims of [deepfake] fabrications can vindicate their rights through tort actions."); Chesney & Citron, *supra* note 20, at 1792–1804.

[59] Va. Code Ann. § 18.2-386.2(A) (2025) (emphasis added).

[60] *Id.*; *see also* 2019 Va. Acts 868 (amending Virginia's revenge porn statute so that it includes deepfakes).

[61] S. 976, 221st Leg., Reg. Sess. (N.J. 2024).

[62] *See* N.J. Stat. Ann. § 2C:14-9(b)(1) (West 2025).

[63] *Id.* (emphasis added).

[64] Pfefferkorn, *supra* note 28, at 253.

[65] *Id.*

[66] Delfino, *supra* note 21, at 1076.

[67] *See* Blitz, *supra* note 17, at 66–67 (discussing the government's potential power to punish false statements in some contexts).

Deepfakes present two challenges to courts' factfinding mission. The first is obvious: Parties may seek to introduce deepfakes as evidence that supports their case.[68] This could be done in a variety of ways:

> [A] video of the crime scene could be manipulated by the perpetrators to change their appearance; an audio recording [could be] manipulated to depict somebody as violent; a criminal could swap their face with somebody else's to create a perfect alibi; [or] an innocent could be framed . . . for revenge.[69]

For example, in one U.K. case, a mother used deepfake audio recordings at a custody hearing to give a false impression that her ex-husband was a danger to their children.[70] Of course, U.S. courts can regulate false content at trial without raising First Amendment concerns.[71] At the time of writing, no reported case in the United States has found that evidence entered into the record was a deepfake.[72] However, one dissenting judge did predict that police could use deepfakes in extrajudicial proceedings to dupe suspects into waiving procedural rights.[73] But the slim official record of deepfakes in evidence does not necessarily mean that deepfakes have yet to infiltrate courtrooms. It only means that they are rarely caught, because "it would never occur to most judges that deepfake material could be submitted as evidence."[74] It is only a matter of time.

Even once judges become aware of the possibility of deepfake evidence, deepfakes would begin to pose a second kind of challenge for courts: The existence of deepfakes can be used at trial to undermine the credibility of legitimate evidence. This has been named the "deepfake defense," and it "is built around the premise that the audiovisual material introduced as evidence against the defendant is claimed

[68] *See* Francesca Palmiotto, Detecting Deep Fake Evidence with Artificial Intelligence 1 (Mar. 11, 2023) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4384122 [https://perma.cc/8SMP-J6AR].

[69] *Id.* at 5.

[70] Reynolds, *supra* note 29.

[71] *See* United States v. Alvarez, 567 U.S. 709, 718–22 (2012) (plurality opinion) (observing that the "unquestioned constitutionality of perjury statutes" arises from the special need for sworn testimony to be reliable so the government may act based on it (quoting United States v. Grayson, 438 U.S. 41, 54 (1978))).

[72] A Westlaw search for "("deepfake" "deep fake")" on all federal and state cases as of September 26, 2024, returns only twenty cases. WESTLAW, + "deepfake" + "deep fake", 20 results (Sep. 26, 2024) (on file with authors).

[73] State v. Garrett, 555 P.3d 1116, 1132 (Kan. 2024) (Rosen, J., dissenting) ("I fear it will not be long before law enforcement tests the limits of creating fabricated images of a detainee at the scene of the crime or artificially create [sic] other evidence in order to convince a suspect to forego their right to remain silent or cooperate with an investigation.").

[74] Patrick Ryan, *'Deepfake' Audio Evidence Used in UK Court to Discredit Dubai Dad*, NATIONAL (Feb. 8, 2020), https://www.thenationalnews.com/uae/courts/deepfake-audio-evidence-used-in-uk-court-to-discredit-dubai-dad-1.975764 [https://perma.cc/HW3N-GNCQ].

to be fake."[75] The deepfake defense has appeared in several reported cases in the United States, though judges presently seem to view it with skepticism.[76] Commentators agree that "the very existence of deepfakes will [inevitably] complicate the task of authenticating *real* evidence."[77]

Apocalyptic concerns voiced by some deepfake scholars extend into the courtroom. They have noted that "[d]eepfakes pose dangers and risks to our society and democratic institutions, including our judicial system."[78] Given the lurking threat of deepfakes, "video evidence may ultimately lose its persuasive power and, if taken far enough, degrade public trust in the very institution of the courts."[79] The fear is that if jurors cannot trust digital media, they may lose their grip on truth itself; and, if there is no truth, what are courts for? "If juries cease believing that the truth exists and that it can be found out, then they will have little cause to keep believing in the courts."[80]

C. Academic Proposals and "Deepfake Detectors"

Even though deepfakes are a relatively recent phenomenon, there is no shortage of proposals for what to do about them. Some commentators have expressed reserved confidence that, at least for the time being, "[w]hen deepfakes result in harm, there are a variety of [existing] laws that may apply to punish and provide restitution."[81] Existing laws could be particularly effective were they applied to the platforms through which many users publish deepfakes.[82] "[B]an[ning] the technology altogether" might provide a longer-term solution,[83] but, as already discussed, "it is unlikely that a flat ban on deep fakes could withstand

[75] Delfino, *supra* note 21, at 1070. The more general phenomenon of undermining media by claiming it could be fake has been dubbed the "liar's dividend." Chesney & Citron, *supra* note 20, at 1758.

[76] *See, e.g.*, People v. Smith, 969 N.W.2d 548, 548, 565 (Mich. Ct. App. 2021) (holding that the trial court did not abuse its discretion in admitting social media posts into evidence that defendant alleged included deepfake photos); *In re* Gabriel H., 215 N.Y.S.3d 613, 617 (N.Y. App. Div. 2024) ("[Respondent] contends that the videos should be given little to no weight because they could be 'deepfakes.' The court afforded the videos great weight based on clear evidence of their reliability . . . ."); Pittman v. Commonwealth, No. 0681-22-1, 2023 WL 3061782, at *6 (Va. Ct. App. Apr. 25, 2023) ("[T]here is no evidence of or contention that would call into question the veracity of the video or the possibility of a 'deep fake.'").

[77] Pfefferkorn, *supra* note 28, at 255.

[78] Delfino, *supra* note 12, at 296.

[79] *Id.* at 312–13.

[80] Pfefferkorn, *supra* note 28, at 276.

[81] Brown, *supra* note 18, at 37.

[82] Chesney & Citron, *supra* note 20, at 1795 ("[T]he most efficient and effective way to mitigate harm may be to impose liability on platforms."). As Chesney and Citron note, section 230 of the Communications Decency Act presently forecloses this possibility. *Id.* at 1795–98; 47 U.S.C. § 230.

[83] Brown, *supra* note 18, at 32.

constitutional challenge."[84] Constitutional law scholars have responded by suggesting more limited bans, such as on deepfakes misrepresenting their "purported source or vehicle," that may have a better chance of passing constitutional muster.[85]

Turning specifically to the courtroom, scholars have proposed modifying procedures and rules of evidence to neutralize deepfakes. One idea for curbing the misuse of the deepfake defense is to amend the rules of procedure to allow courts to sanction attorneys who in bad faith question the authenticity of digital media during oral argument.[86] Most of the attention, however, has been devoted to the challenge of excluding deepfakes from evidence. Here, the authentication standard embodied in rules like Federal Rule of Evidence 901 plays a prominent role.[87] One scholar would "expand the gatekeeping function of the court by assigning the responsibility of deciding authenticity issues [for digital media] solely to the judge."[88] Another would require "a person whose occupation or means of knowledge is in a specialized field . . . to testify about the [digital] evidence" where there are questions of authenticity.[89] A final proposal would ask "[j]udges and attorneys . . . to find the originator of a video or photo" and warns that "it may no longer be prudent to admit video evidence when the origin of a video is indeterminable."[90] Widespread adoption of self-certifying technology on media-capture devices could augment this approach.[91]

Some evidence scholars believe that technological developments will save the day without a need to amend current law.[92] As they observe, there is a "long history of fakery" in evidence, even where digital images were concerned.[93] Since the release of Photoshop in 1990, users have

[84] Chesney & Citron, *supra* note 20, at 1790.

[85] Blitz, *supra* note 17, at 63–64 ("Fake news may lose protection, I suggest, when it is not only a falsity, but a forgery as well.").

[86] Delfino, *supra* note 21, at 1071. Currently, Federal Rule of Civil Procedure 11 only applies to signed writings, *id.* at 1092, and only in civil trials, *id.* at 1095.

[87] McPeak, *supra* note 22, at 440 ("[P]roper use of the authentication rules in the Federal Rules of Evidence can alleviate both concerns [with deepfakes]."); *see* Fed. R. Evid. 901 ("[T]o satisfy the requirement of authenticating . . . evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is.").

[88] Delfino, *supra* note 12, at 341.

[89] Molly Mullen, *A New Reality: Deepfake Technology and the World Around Us*, 48 Mitchell Hamline L. Rev. 210, 229 (2022).

[90] John Channing Ruff, Note, *The Federal Rules of Evidence Are Prepared for Deepfakes. Are You?*, 41 Rev. Litig. 103, 123 (2021).

[91] Delfino, *supra* note 12, at 341 ("As self-authenticating software becomes available on more devices, a court may be able to look to Rule 902(13) and (14) to make the required determination of authenticity.").

[92] *See* Pfefferkorn, *supra* note 28, at 266 ("[C]ourts are confident in the processes they already have in place for excluding manipulated evidence. I share that confidence.").

[93] *See id.* at 256.

been able to alter digital images.[94] But "[n]o major regulation or legislation was needed to prevent the apocalyptic vision of Photoshop's future; society adapted on its own."[95] In the same way, courts may develop "strategies for keeping deepfake videos out of evidence,"[96] relying on experts where needed[97] or using their own "training in spotting outward signs of altered deepfake technology."[98] Because "deepfakes are still generally not very good," significantly altering the Federal Rules of Evidence at this stage would be a "gross overreaction."[99]

All the scholars discussed in this Section seem to assume that there is—and always will be—some way to distinguish deepfakes from genuine media. Banning deepfake content or asking platforms to police it can only be effective if there is some reliable way to tell when media are fake. For confronting deepfakes at trial, these scholars note that "[i]dentifying potentially deepfake content is just the first of the necessary steps" in the solutions they envision.[100] Evidence scholars recommend "a 'go-slow-and-strict' approach . . . to allow for the development of better technologies that can detect deepfakes."[101] In other words, "[deepfake] detectors [will be] indispensable for every party to the criminal process."[102] As discussed next, technologists are not so optimistic about the long-term prospects for such detectors.

D. The Enduring Challenge of Deepest Fakes

There is an active research community focused on developing sophisticated techniques for detecting deepfakes. They have found several tells. Sometimes there are digital artifacts, which include unintended products of the technology used to make deepfakes.[103] For example, there may be a discrepancy between the expected file size of a

[94] Adobe Corp. Commc'ns, *Adobe Explains It All: Photoshop*, ADOBE BLOG (Feb. 25, 2015), https://blog.adobe.com/en/publish/2015/02/25/adobe-explains-it-all-photoshop [https://perma.cc/2WN7-JTT4].

[95] Jeffrey Westling, *Deep Fakes: Let's Not Go off the Deep End*, TECHDIRT (Jan. 30, 2019, at 12:05 ET), https://www.techdirt.com/articles/20190128/13215341478/deep-fakes-lets-not-go-off-deep-end.shtml [https://perma.cc/M457-8MNY].

[96] Pfefferkorn, *supra* note 28, at 259.

[97] *Id.* at 262–63.

[98] Mullen, *supra* note 89, at 224.

[99] Ruff, *supra* note 90, at 125.

[100] Brown, *supra* note 18, at 58.

[101] Delfino, *supra* note 12, at 316.

[102] Palmiotto, *supra* note 68, at 9 ("[I]t is crucial to make [deepfake detectors] available to all the involved parties.").

[103] *See* John Spacey, *7 Types of Data Artifact*, SIMPLICABLE (Apr. 20, 2017), https://simplicable.com/new/data-artifact [https://perma.cc/2DQT-BU4Q].

video and its actual size,[104] or there may be subtle clues left by "'digital manipulations such as scaling, rotation or splicing' that are commonly employed in deepfakes."[105] Researchers at the Massachusetts Institute of Technology and the Department of Defense have taken a different approach, examining subtle biometric markers that deepfakes sometimes botch, like distorting microdetails of a person's iris or failing to match a person's blood pulse in all parts of their body.[106] Other biomarkers include rates of eye blinking[107] or small distortions in facial regions.[108]

There is just one problem with this approach: Every new deepfake detector ultimately helps refine deepfake generators. Deepfake generators are made using generative adversarial networks.[109] These "are two-part AI models consisting of a generator that creates samples [of video, images, or audio] and a discriminator that attempts to differentiate between the generated samples and real-world samples."[110] Each success of the discriminator feeds back into the training of the generator, continually improving its ability to produce realistic outputs.[111] For example, "The same deep-learning technique that can spot face-swap videos can also be used to improve the quality of face swaps in the first

[104] Kaveh Waddell, *The Impending War over Deepfakes*, Axios (July 22, 2018), https://www.axios.com/2018/07/22/the-impending-war-over-deepfakes [https://perma.cc/58RY-FHKA].

[105] John Villasenor, *Artificial Intelligence, Deepfakes, and the Uncertain Future of Truth*, Brookings Inst. (Feb. 14, 2019) (quoting Jason Bunk et al., Detection and Localization of Image Forgeries Using Resampling Features and Deep Learning 1 (July 3, 2017) (unpublished manuscript), https://arxiv.org/pdf/1707.00433 [https://perma.cc/RBH4-VE4L]), https://www.brookings.edu/blog/techtank/2019/02/14/artificial-intelligence-deepfakes-and-the-uncertain-future-of-truth/ [https://perma.cc/GFC9-7FMH].

[106] *See* Satya Venneti, *Real-Time Extraction of Biometric Data from Video*, Carnegie Mellon Univ. Software Eng'g Inst.: Blog (Aug. 22, 2016), https://insights.sei.cmu.edu/blog/real-time-extraction-of-biometric-data-from-video/ [https://perma.cc/BA9W-UFEX]; Peipeng Yu, Zhihua Xia, Jianwei Fei & Yujiang Lu, *A Survey on Deepfake Video Detection*, 10 IET Biometrics 607, 616 (2021).

[107] Yuezun Li, Ming-Ching Chang & Siwei Lyu, In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking 1 (June 11, 2018) (unpublished manuscript), https://arxiv.org/pdf/1806.02877 [https://perma.cc/MJD3-YG8G].

[108] Palmiotto, *supra* note 68, at 8; *see* Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies & Matthias Nießner, FaceForensics: A Large-Scale Video Dataset for Forgery Detection in Human Faces 1 (Mar. 24, 2018) (unpublished manuscript), https://arxiv.org/pdf/1803.09179 [https://perma.cc/64W7-CAP5]; Emerging Technology from the arXiv, *This Algorithm Automatically Spots "Face Swaps" in Videos*, MIT Tech. Rev. (Apr. 10, 2018), www.technologyreview.com/s/610784/this-algorithm-automatically-spots-face-swaps-in-videos/ [https://perma.cc/K5P7-625Y].

[109] Ian J. Goodfellow et al., Generative Adversarial Nets 1 (June 10, 2014) (unpublished manuscript), https://arxiv.org/pdf/1406.2661 [https://perma.cc/MC9E-79WD].

[110] Kyle Wiggers, *Generative Adversarial Networks: What GANs Are and How They've Evolved*, VentureBeat (Dec. 26, 2019) (emphasis omitted), https://venturebeat.com/2019/12/26/gan-generative-adversarial-network-explainer-ai-machine-learning/ [https://perma.cc/VQS7-TSKL].

[111] *Id.*

place—and that could make them harder to detect."[112] Indeed, shortly after researchers discovered the eye-blinking test as a tell for detecting deepfakes, deepfake generators had figured out how to defeat it.[113]

It is little surprise, then, that "leading digital forensics experts worry that the fight to detect deepfakes is a losing battle—that deepfake technology is outstripping the ability of those trying to detect the deepfakes."[114] It is a game of cat and mouse; and, as in the Sunday-morning cartoons, the mouse is always wilier. Research on deepfake detection will continue, but each new advancement requires creative effort and innovation. Deepfake generators, by contrast, only need one trick to adapt: retrain the model using the new detector's output as data.[115] Laypeople cannot tell today's deepfakes apart from genuine media,[116] and even experts are having trouble.[117] As scholars have noted, "a variety of detection mechanisms exist, and they are improving. But they still lag behind the sophistication of deepfakes, which continue to advance."[118] Technologists put it more bluntly: "The adversary will always win . . . ."[119]

[112] Emerging Technology from the arXiv, *supra* note 108.

[113] John P. LaMonaca, *A Break from Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes*, 69 AM. U. L. REV. 1945, 1956–57 (2020).

[114] Delfino, *supra* note 12, at 346; *see* Drew Harwell, *Top AI Researchers Race to Detect 'Deepfake' Videos: 'We Are Outgunned,'* WASH. POST (June 12, 2019), https://www.washingtonpost.com/technology/2019/06/12/top-ai-researchers-race-detect-deepfake-videos-we-are-outgunned/ [https://perma.cc/RFB3-4VDF]; Hilke Schellmann, *The Dangerous New Technology That Will Make Us Question Our Basic Idea of Reality*, QUARTZ (July 20, 2022), https://qz.com/1145657/the-dangerous-new-technology-that-will-make-us-question-our-basic-idea-of-reality [https://perma.cc/23V9-DMJN] ("[F]orensic specialists predict that computers will be able to generate convincing, fabricated audio and video recordings at a rapid pace in the next few years.").

[115] Louise Matsakis, *Artificial Intelligence Is Now Fighting Fake Porn*, WIRED (Feb. 14, 2018, at 16:46 ET), https://www.wired.com/story/gfycat-artificial-intelligence-deepfakes [https://perma.cc/W59K-JTC8] ("If you really want to fool the system you will start building into the deepfake ways to break the forensic system.").

[116] Nils C. Köbis, Barbora Doležalová & Ivan Soraperra, *Fooled Twice: People Cannot Detect Deepfakes but Think They Can*, ISCIENCE 1 (Nov. 19, 2021), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8602050/pdf/main.pdf [https://perma.cc/W23L-9TJX].

[117] Thanh Thi Nguyen et al., Deep Learning for Deepfakes Creation and Detection: A Survey 13 (Aug. 11, 2022) (unpublished manuscript), https://arxiv.org/pdf/1909.11573 [https://perma.cc/Y7A9-ZJ4R].

[118] *See* Brown, *supra* note 18, at 23; *see also* Harris, *supra* note 11, at 102 ("[T]his type of production carries immense potential to be indistinguishable from real-life videos."); Dan Boneh, Andrew J. Grotto, Patrick McDaniel & Nicolas Papernot, How Relevant Is the Turing Test in the Age of Sophisbots? 3 (Aug. 30, 2019) (unpublished manuscript), https://arxiv.org/pdf/1909.00056 [https://perma.cc/JTD9-RX2W] ("[I]n the long-run [deepfake detection] is likely to be a losing battle or at best a stalemate.").

[119] Dan Robitzski, *DARPA Spent $68 Million on Technology to Spot Deepfakes*, FUTURISM (Nov. 19, 2018, at 14:35 ET) (quoting Stephanie Kampf & Mark Kelley, *A New 'Arms Race': How the U.S. Military Is Spending Millions to Fight Fake Images*, CANADIAN BROAD. CORP. (Nov. 18, 2018), https://www.cbc.ca/news/science/fighting-fake-images-military-1.4905775 [https://perma.

A more promising technical approach may be to mechanically authenticate media as unaltered rather than attempting to expose deepfakes. For example, location verification "is available already, thanks to the ubiquity of phones with location tracking features as well as cell-site location records."[120] It could help demonstrate that the camera and the subject were in the same place at the same time.[121] Additionally, some media-capture devices now come equipped with cryptographic or similar authentication processes, which could be used to verify that the media originated from the device in question.[122]

Unfortunately, even sophisticated authentication is far from foolproof. Location verification is ineffective if global positioning system locations can be spoofed—a vulnerability that has existed for decades.[123] And device authentication provides little assurance when media can be altered on the device. In 2018, researchers showed how to do this with police bodycam footage.[124] Even if there were methods to defeat location spoofing and on-device manipulation, there is a simple workaround: Adversaries could use their media-capture devices *in situ* to record the output of a second device playing a deepfake.

Five years ago, Bobby Chesney and Danielle Citron warned of "a worst-case scenario . . . in which it is cheap and easy to [make deepfakes] . . . with inadequate technology to quickly and reliably expose [them]."[125] We anticipate an even worse worst-case scenario in which cheap and easy deepfakes are immune not only to quick and reliable detection but to any detection at all. Deepfake generators will eventually evolve to create what this Article calls "deepest fakes"—costless deepfakes that no procedure relying on the content of the media alone can distinguish from authentic media. Deepest fakes should shake the confidence of scholars who think that expert testimony and detection techniques will rescue courts from the coming upheaval. Deepest fakes

cc/6ZPH-KW2H]), https://futurism.com/darpa-68-million-technology-deepfakes [https://perma.cc/XX24-G8XX].

[120] Chesney & Citron, *supra* note 20, at 1815.

[121] *See id.*

[122] *See, e.g.*, Lily Hay Newman, *Police Bodycams Can Be Hacked to Doctor Footage*, WIRED (Aug. 11, 2018, at 15:00 ET), https://www.wired.com/story/police-body-camera-vulnerabilities/ [https://perma.cc/CV4R-3BGH].

[123] *See* Nils Ole Tippenhauer, Christina Pöpper, Kasper B. Rasmussen & Srdjan Čapkun, *On the Requirements for Successful GPS Spoofing Attacks*, *in* CCS'11: PROCEEDINGS OF THE 18TH ACM CONFERENCE ON COMPUTER & COMMUNICATIONS SECURITY 75, 75 (2011). Today, even children use spoofing to access new locations in augmented-reality videogames. *See, e.g.*, *A Pokémon Go Tool That Takes You Everywhere!*, SPOOFER GO, https://www.spoofer-go.com/ [https://perma.cc/J8WJ-SLCA] (last visited Oct. 21, 2025) ("Supported by Pokémon Go, Spoofer Go has a powerful fake location function along with great movement functions that make exploring the Pokémon Go world easier and more exciting.").

[124] Newman, *supra* note 122.

[125] Chesney & Citron, *supra* note 20, at 1814.

undermine all existing proposals for devising enhanced authentication requirements for digital media. In a world in which deepest fakes are prevalent, the only distinguishing feature of inauthentic media would be that it portrays an event that never occurred.[126] Verifying authenticity would create an impossible circularity—demanding proof of the event that the media itself serves to prove.

II. Epistemology and Deepest Fakes

The threat of deepest fakes may be recent, but the general worry is hardly novel. In his 1641 work *Meditations on First Philosophy*, René Descartes imagines that everything he thought he knew is merely an illusion:

> I will suppose therefore that . . . some malicious demon of the utmost power and cunning has employed all his energies in order to deceive me. I shall think that the sky, the air, the earth, colours, shapes, sounds and all external things are merely the delusions of dreams which he has devised to ensnare my judgment. I shall consider myself as not having hands or eyes, or flesh, or blood or senses, but as falsely believing that I have all these things. . . . [E]ven if it is not in my power to know any truth, I shall at least . . . resolutely guard against assenting to any falsehoods . . . .[127]

Descartes then asks whether any kind of knowledge is possible in the face of such extreme doubt.[128]

Deepest-fake generators are a far cry from Descartes's demon, but they do similarly destabilize an information environment we formerly trusted. Descartes's thought experiment forces us to confront the possibility of skepticism about perception. Can we trust our eyes and ears to deliver true information about reality, or, as Edgar Allan Poe writes, could "[a]ll that we see or seem [be] but a dream within a dream"?[129] Even if our eyes and ears *could* deliver true information about reality, the very possibility that Descartes's demon *might* be manipulating what we see and hear undermines the extent to which we can trust perception. The eventuality of deepest fakes forces a parallel skeptical worry:

[126] Blitz, *supra* note 17, at 68 ("[F]alse factual statements are unlike religious ideas and political opinions in at least one respect: they can be exposed as fake.").

[127] René Descartes, Meditations on First Philosophy 19 (John Cottingham ed. & trans., Cambridge Univ. Press 2d ed. 2017) (1641).

[128] *Id.*

[129] Edgar A. Poe, *A Dream Within a Dream*, Flag Our Union (Bos.), Mar. 31, 1849, at 2 (emphasis omitted) (on file with Univ. of Tex. Aus., Harry Ransom Center, Edgar Allen Poe Collection, https://hrc.contentdm.oclc.org/digital/collection/p15878coll102/id/2883/rec/23 [https://perma.cc/UXE9-4KMF]).

In the age of deepest fakes, could we ever rely on digital media, or must we treat it all as the digital equivalent of hallucination? This is the question that motivates deepfake alarmism.

This Part provides philosophical tools that help to structure a response. It begins by introducing some basic concepts from epistemology, including skepticism.[130] Although antiskeptical philosophers have many responses to Descartes's thought experiment, these responses are surprisingly ineffective at addressing deepfake alarmism. As this Part shows, deepfake alarmism poses a philosophical challenge that is in some respects less tractable than even Cartesian skepticism. Although there is a nascent philosophical literature trying to characterize and ameliorate the epistemic threat that deepfakes pose, their solutions misfire when confronted with deepest fakes or the particular challenges of the courtroom context.[131] All is not lost, though. There are elements of epistemologists' views[132] that, modified and extended as we propose below, motivate strategies for courts[133] and ordinary people[134] to adapt in an increasingly unreliable digital landscape.

A. *Digital Media Skepticism*

Epistemology is the philosophical study of knowledge and belief.[135] It offers a conceptual framework for understanding the ways that deepfakes undermine our ability to know about the world around us. More important, epistemologists elucidate alternate pathways to knowledge.

According to the classic definition,[136] "knowledge" is justified true belief.[137] "Belief" is a mental state that represents reality as being a particular way: For example, "It is raining outside."[138] A belief is "true" if

[130] *See infra* Section II.A.

[131] *See infra* Section II.B.

[132] *See infra* Section II.C.

[133] *See infra* Part IV.

[134] *See infra* Conclusion.

[135] *Epistemology*, Merriam-Webster, https://www.merriam-webster.com/dictionary/epistemology [https://perma.cc/9QU4-SQNM] (last visited Oct. 21, 2025).

[136] The discussion throughout this Article focuses only on *a posteriori* knowledge—i.e., knowledge about the outside world obtained through experience. *See* Bruce Russell, *A Priori Justification and Knowledge*, *in* Stanford Encyclopedia of Philosophy § 3 (Edward N. Zalta & Uri Nodelman eds., summer 2024 ed. 2024), https://plato.stanford.edu/archives/sum2024/entries/apriori/ [https://perma.cc/A56L-Q66C]. *A priori* knowledge—i.e., knowledge obtained through reason alone—is not relevant here. *See id.*

[137] Paul K. Moser & Arnold vander Nat, Human Knowledge 12–15 (3d ed. 2003). Note, this is different from how the legal system often characterizes knowledge, which is merely as true belief. Mihailis E. Diamantis, *Functional Corporate Knowledge*, 61 Wm. & Mary L. Rev. 319, 334–35 (2019). Most philosophers today add various other complicating requirements that need not detain us here. *See* Edmund L. Gettier, *Is Justified True Belief Knowledge?*, 23 Analysis 121, 121 (1963).

[138] *See* David Hume, A Treatise of Human Nature 624 (L.A. Selby-Bigge ed., Oxford, Clarendon Press 1896) (1739).

it represents reality accurately, because, for example, it *really is* raining outside.[139]

The justification element in the definition of knowledge is much more contested.[140] Two dominant views on epistemic justification are evidentialism and reliabilism. Evidentialists maintain that a person is justified in believing a proposition if that person possesses sufficient evidence that it is true.[141] Types of evidence include perception (you see the rain outside your window), introspection (you experience the pain in your knee that often precedes rain), memory (you remember seeing rain clouds rolling in this morning), intuition (you had a premonition of rain), and testimony (your local meteorologist tells you it is raining).[142] Depending on the circumstance, some types of evidence will be stronger than others. Seeing that it is raining is usually enough to justify the belief that it is raining, but merely hearing a pitter-patter on the roof may need supplementing by other evidence. Evidentialists usually do not specify a bright-line threshold for what counts as sufficient evidence, and the threshold may vary.[143] For example, more evidence may be needed for forming justified beliefs about matters of consequence (whether it is safe to go sailing today) than for relative trivialities (whether you will need to mow your grass again next week).[144]

Reliabilism is the view that beliefs are justified if they are formed using a process that usually results in true beliefs.[145] For example, someone could form the belief that it will rain tomorrow either by checking the weather forecast or their daily horoscope. The process that involves the forecast is reliable and would result in a justified belief. Not so for the horoscope. Some processes (e.g., normal human vision in good lighting) are unconditionally reliable, delivering appropriate outputs in most situations, while others (e.g., deductive inference) are only conditionally reliable, depending on correct inputs. Like evidentialists, reliabilists usually do not specify a precise threshold for how reliable—in other words,

[139] MOSER & NAT, *supra* note 137, at 9–11.

[140] *See infra* notes 141–46 and accompanying text.

[141] *See* Jaegwon Kim, *What Is "Naturalized Epistemology?," in* 2 PHILOSOPHICAL PERSPECTIVES: EPISTEMOLOGY 381, 390–91 (James E. Tomberlin ed., 1988); Richard Feldman & Earl Conee, *Evidentialism*, 48 PHIL. STUD. 15, 15 (1985).

[142] *See* Jeremy Fantl & Matthew McGrath, *Evidence, Pragmatics, and Justification*, 111 PHIL. REV. 67, 85 (2002).

[143] JASON STANLEY, KNOWLEDGE AND PRACTICAL INTERESTS 1–6 (2005).

[144] *Id.*; Fantl & McGrath, *supra* note 142, at 85.

[145] FRANK PLUMPTON RAMSEY, *Knowledge*, *in* THE FOUNDATIONS OF MATHEMATICS 258, 258 (R.B. Braithwaite ed., 1931); ALVIN I. GOLDMAN, EPISTEMOLOGY AND COGNITION 108 (1986). There are other variants of reliabilism, *see, e.g.*, ERNEST SOSA, KNOWLEDGE IN PERSPECTIVE 131 (1991) (discussing virtue reliabilism); Peter Unger, *An Analysis of Factual Knowledge*, 65 J. PHIL. 157, 159 (1968) (defending a version of reliabilism according to which someone is justified in holding a belief if and only if "it is not at all accidental" that the belief is true), but the process-oriented view is the most common.

how likely to result in true beliefs—a belief-forming process must be and acknowledge that the threshold may depend on context.[146]

Often, reliabilists and evidentialists give the same answers in ordinary cases.[147] This is to be expected because both aspire to reflect commonsense intuitions about when beliefs are justified and, hence, candidates for being knowledge. For example, both views would hold that a person's belief that it is raining is justified if that is what the meteorologist told them—such a person both possesses sufficient evidence and formed their belief using a reliable process. The views may give different results in some exotic cases involving wishful thinking,[148] alternate universes,[149] or clairvoyance.[150] We will not attempt to referee which view is superior. For present purposes, it suffices to note that evidentialism and reliabilism offer different conceptions of epistemic justification, one focused on evidence and the other focused on process.

Skepticism is the worry that we are not actually justified in believing some things we think we know.[151] Different varieties of skepticism target different classes of beliefs.[152] Other-minds skepticism, for example, arises from the solipsistic worry that everyone else might be unconscious.[153] It aims to undermine the justification for our beliefs about other people's minds.[154] Pyrrhonian skepticism, by contrast, calls all knowledge into doubt by questioning whether we are ever justified in believing anything.[155] Cartesian skepticism falls somewhere between the two. Recall that Descartes entertains the thought that "the sky, the

[146] *See* Alvin Goldman & Bob Beddor, *Reliabilist Epistemology*, *in* Stanford Encyclopedia of Philosophy § 1.1 (Edward N. Zalta ed., summer 2021 ed. 2021), https://plato.stanford.edu/archives/sum2021/entries/reliabilism/ [https://perma.cc/FU6K-ABB6] ("Precisely how high a truth-ratio must be in order to confer justifiedness is left vague . . . .").

[147] Indeed, some epistemologists have argued for a unified approach. *See, e.g.*, Juan Comesaña, *Evidentialist Reliabilism*, 44 Noûs 571, 571 (2010); Alvin I. Goldman, *Toward a Synthesis of Reliabilism and Evidentialism? Or: Evidentialism's Troubles, Reliabilism's Rescue Package*, *in* Evidentialism and Its Discontents 254, 254 (Trent Dougherty ed., 2011).

[148] *See* Alvin I. Goldman, *What Is Justified Belief?*, *in* Justification and Knowledge 1, 16–18 (George S. Pappas ed., 1979).

[149] *See* Hilary Putnam, *The Meaning of "Meaning," in* 7 Minnesota Studies in the Philosophy of Science 131, 148–51 (Keith Gunderson ed., 1975); Tyler Burge, *Individualism and the Mental*, 4 Midwest Stud. Phil. 73, 77–79 (1979).

[150] *See* Laurence Bonjour, *Externalist Theories of Empirical Knowledge*, 5 Midwest Stud. Phil. 53, 59–60 (1980); Ralph Wedgwood, *The Aim of Belief*, 16 Phil. Persps. 267, 278–79 (2002).

[151] Peter Klein, *Skepticism*, *in* Stanford Encyclopedia of Philosophy § 1 (Edward N. Zalta ed., summer 2015 ed. 2015), https://plato.stanford.edu/archives/sum2015/entries/skepticism/ [https://perma.cc/JYR9-S39P].

[152] James Conant & Andrea Kern, *Introduction: From Kant to Cavell*, *in* Varieties of Skepticism 1, 1–2 (James Conant & Andrea Kern eds., 2014).

[153] *See* Anita Avramides, Other Minds 1–4 (2001).

[154] *See id.*

[155] *See* Juan Comesaña, *Pyrrhonian Problematic, The*, *in* 1 Encyclopedia of Philosophy 174, 174–75 (Donald M. Borchert ed., 2d ed. 2006).

air, the earth, colours, shapes, sounds and all external things are merely the delusions of [our] dreams."[156] What he articulates is external-world skepticism: the view that we cannot know that there is an external world or know anything about it.[157] Although Descartes's thought experiment involves us being manipulated by an evil demon, more modern variants raise the possibility that we might just be brains in vats[158] or figments of some vast computer simulation.[159]

Both evidentialism and reliabilism must confront Descartes's radical doubt. If such widespread demonic deception is possible, how can anyone be confident in their evidence or in the reliability of their belief-forming processes? We ordinarily take perception to provide sufficient evidence for beliefs about the external world.[160] That presumption assumes that there are no genuine grounds for doubting our perception.[161] If we suspect we have ingested a hallucinogenic drug or are being manipulated by an evil demon, then grounds for doubt start to creep in, and perception may lose its justificatory power.

Descartes's strategy for overcoming external-world skepticism relied on an elaborate argument against the very possibility of the evil demon. He started by identifying at least one thing he could know with certainty, even in the grips of a demon's illusions:

> [L]et [the demon] deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something. So after considering everything very thoroughly, I must finally conclude that this proposition, *I am*, *I exist*, is necessarily true whenever it is put forward by me or conceived in my mind.[162]

This is the famous "cogito" argument: "*I think therefore I am*."[163] Although the cogito has some undeniable plausibility, fewer people find the rest of Descartes's argument against skepticism very persuasive. He leverages the cogito to conclude that there is an all-powerful God, that

[156] DESCARTES, *supra* note 127, at 19; *see also* Lex Newman, *Descartes on the Method of Analysis* (discussing Descartes's notion that all analysis requires certain assumptions about reality), *in* THE OXFORD HANDBOOK OF DESCARTES AND CARTESIANISM 65, 72–73 (Steven Nadler et al. eds., 2019).

[157] *See* Newman, *supra* note 156, at 72–73.

[158] HILARY PUTNAM, REASON, TRUTH AND HISTORY 2, 5–6 (1981); David J. Chalmers, *The Matrix as Metaphysics*, *in* PHILOSOPHERS EXPLORE THE MATRIX 132, 132 (Christopher Grau ed., 2005).

[159] Nick Bostrom, *Are We Living in a Computer Simulation?*, 53 PHIL. Q. 243, 243 (2003).

[160] Ali Hasan, *The Evidence in Perception*, *in* THE ROUTLEDGE HANDBOOK OF THE PHILOSOPHY OF EVIDENCE 292, 293 (Maria Lasonen-Aarnio & Clayton Littlejohn eds., 2024).

[161] Klein, *supra* note 151, § 1.

[162] DESCARTES, *supra* note 127, at 21.

[163] *See* Lex Newman, *Descartes' Epistemology*, *in* STANFORD ENCYCLOPEDIA OF PHILOSOPHY § 4 (Edward N. Zalta & Uri Nodelman eds., winter 2023 ed. 2023), https://plato.stanford.edu/archives/win2023/entries/descartes-epistemology/ [https://perma.cc/BF2Q-7SKL].

God must be benevolent, and that a benevolent God would not permit an evil demon to deceive us.[164]

Modern evidentialists take a different approach to external-world skepticism, offering a multipronged response that avoids Descartes's reliance on metaphysical guarantees. According to one prominent evidentialist position, "dogmatism," our perceptual beliefs about the external world have prima facie justificatory power for beliefs about the circumstances they convey.[165] This justification stands so long as there is no specific evidence that undermines or contradicts our perceptual evidence—i.e., so long as there are no "defeaters."[166] Importantly, there is no evidence for skeptical hypotheses like the evil demon. For the dogmatist, the mere conceptual possibility of an evil demon does not defeat the prima facie power of perceptual justification.[167]

Evidentialists have a second type of response to Cartesian skepticism: inference to the best explanation.[168] In short, compared to skeptical alternatives like the evil demon, the hypothesis that there is an external world more simply and coherently explains several attributes of our perceptual experience: Perceptual experiences tend to be stable over time (stop signs look red on every day of the week), people tend to perceive things similarly (stop signs look red to nearly all of us), and perceptions generally facilitate practical success (people who stop when they see stop signs live longer).[169] If there were an evil demon, we would need a rather convoluted story to explain why it cares to assure our perceptual experiences have these traits.

Unlike evidentialists, reliabilists respond to external-world skepticism by examining the reliability of the belief-forming process.[170] A common reliabilist response invokes the concept of epistemic safety.[171] A belief is considered safe only if a person could not easily come to hold it without it being true.[172] "Easily" is doing a lot of work in this definition. There is a technical definition of "easily" using possible-world semantics,[173] but the general idea turns on how radically different the

[164] *See* Descartes, *supra* note 127, at 17–18.

[165] James Pryor, *The Skeptic and the Dogmatist*, 34 Noûs 517, 518 (2000).

[166] Hasan, *supra* note 160, at 293.

[167] Fred I. Dretske, *Epistemic Operators*, 67 J. Phil. 1007, 1011–12 (1970); G.C. Stine, *Skepticism, Relevant Alternatives, and Deductive Closure*, 29 Phil. Stud. 249, 253–54 (1976). Of course, the evidentiary landscape would be very different if the evil demon revealed itself to our perceptions.

[168] Jonathan Vogel, *Cartesian Skepticism and Inference to the Best Explanation*, 87 J. Phil. 658, 658 (1990).

[169] *See id.*

[170] Richard Feldman, *Reliability and Justification*, 68 Monist 159, 159 (1985); Richard Foley, *What's Wrong with Reliabilism?*, 68 Monist 188, 189 (1985).

[171] Ernest Sosa, *Tracking, Competence, and Knowledge*, *in* The Oxford Handbook of Epistemology 264, 274 (Paul K. Moser ed., 2005).

[172] Duncan Pritchard, *Anti-Luck Epistemology*, 158 Synthese 277, 281 (2007).

[173] *See id.* at 282–84.

world would have to be for someone to form a belief based on the same evidence without the belief being true.[174] The safety condition enriches the reliabilist conception of reliability by adding the requirement of robustness across possible worlds: Not only must the process that produces a belief be reliable in the actual world, but the belief must remain true across nearby changes in circumstance.[175]

An example will help. Suppose Tina believes her favorite band is playing because that is what she thinks she sees and hears. She also believes that she is at a rave and that hallucinogens are plentiful at raves. Her belief that her favorite band is playing is probably not very safe. Her perceptions *might* be distorted by a drug. It is not such a remote possibility that she accidentally ingested a hallucinogen or that she purposely took it and forgot. Under the effects of the hallucinogen, Tina might mistake a second-rate cover band for her favorite band. In other words, she might easily have the same belief without the belief being true. Applying the safety condition, reliabilists could conclude that Tina's belief does not amount to knowledge, even if her favorite band actually is playing.

Contrast Tina's belief about her favorite band with our belief in an external world. We move about in what we take to be a real external world filled with coffee shops, honking cars, and street musicians. We believe all these things are real because that is what we think we see and hear. Of course, as Descartes observed, it is *possible* that we are in the grips of an extended illusion orchestrated by an evil demon.[176] But that possible scenario is very remote from the world we believe we inhabit—many things would have to be very different. It follows that our belief in an external world is safe, and Descartes's hypothetical does not undermine our knowledge.

What has all of this got to do with deepfakes? One way to understand the position of deepfake alarmists is that they are positing a new variety of skepticism. We might call it "digital media skepticism." Although we ordinarily take ourselves to be justified in believing something after we have seen a video recording of it, digital media skepticism claims that the possibility that the recording could be a deepfake undermines our justification; we "may doubt [even] unaltered content simply because [we] know realistic deepfakes are possible."[177] The skeptical worry only grows as deepfakes become more prevalent and more realistic—i.e., as deepest fakes come into the picture.

Surprisingly, digital media skepticism appears to be even more philosophically intractable than external-world skepticism. Under the conditions that concern this Article—adversarial trial in a nearby future

[174] *See id.* at 281–82.

[175] *See id.* at 282–84.

[176] DESCARTES, *supra* note 127, at 19.

[177] Delfino, *supra* note 12, at 337.

in which deepest fakes exist—the standard evidentialist and reliabilist solutions to external-world skepticism do not work.

For evidentialists, deepest fakes make it much harder for digital media to provide justifying evidence. Dogmatism about the justificatory power of digital media fails because evidence that deepfakes exist defeats the prima facie credibility of all digital media. The risk of a deepest fake is highly pertinent, not remote. Indeed, there are digital media that tell us about the threat of deepfakes.[178] So, paradoxically, trusting digital media requires us to distrust digital media. Inference to the best explanation also struggles to justify trust in digital media because there are easy explanations for why convincing forgeries exist.

For reliabilists, digital media skepticism is more troubling than external-world skepticism precisely because it is not some far-fetched scenario. Although ordinary beliefs about the external world based on direct perception may be safe, beliefs based solely on digital media are decidedly unsafe.[179] Digital media could very easily convey false content because that is what deepest fakes do. Forming beliefs on the basis of digital media becomes a much less reliable process, particularly when highly motivated courtroom adversaries provide them. As deepest fakes become increasingly prevalent, there will be no sensibly "normal" reference context for reassessing the reliability of the process of forming beliefs using digital media.

Digital media skepticism raises an additional unsettling worry for both evidentialists and reliabilists that standard presentations of external-world skepticism do not. On typical formulations of the evil-demon hypothetical, the demon induces perceptions that reflect the sorts of experiences we ordinarily have. This means it looks to us like there is an external world, and part of the explanation is that our perceptions are stable, composed, and consistent. Digital media skepticism, by contrast, envisions a world in which digital media simultaneously present directly conflicting representations of the same reality. If the conflicting videos happen to be deepest fakes, there will be no internal indication that either video is more reliable or provides better evidence. This raises to salience the fact that at least one of the videos misrepresents reality—and possibly both do!

B. Backstops, Signals, and Norms

There is a nascent philosophical literature about deepfakes. Ethicists focus on the moral harm that deepfakes can cause. Adrienne de

[178] *See supra* notes 1–10 and accompanying text.

[179] *See* Taylor Matthews & Ian James Kidd, *The Ethics and Epistemology of Deepfakes*, *in* THE ROUTLEDGE HANDBOOK OF PHILOSOPHY AND MEDIA ETHICS 342, 344–46 (Carl Fox & Joe Saunders eds., 2024).

Ruiter, for example, uses Kantian ethics to characterize nonconsensual deepfakes as a form of "digital persona plagiarism."[180] Epistemologists aim to characterize the "epistemic harm" of deepfakes.[181] Pessimists envision a future in which deepfakes induce widespread digital media skepticism, and digital media skepticism undermines other knowledge practices. Optimists think they have identified systems that could mitigate deepfakes' epistemic harms. Both groups tend to consider the impact that existing deepfake technology will have on people as they go about their ordinary lives. Deepest fakes in the courtroom raise a novel set of challenges that amplify pessimists' concerns[182] and neutralize optimists' solutions.[183]

1. The Epistemic Harm of Deepfakes

Don Fallis and Regina Rini offer the two most influential philosophical accounts of the ways that deepfakes can distort or undermine our ability to form true and justified beliefs. Fallis focuses on the beliefs we form using digital media.[184] He argues that such beliefs have become integral to how we learn about the world.[185] Although direct perception may be the evidentiary gold standard, "we cannot always be at the right place, at the right time, to see things for ourselves. In such cases, videos are often the next best thing."[186] It is one thing to read about the devastation in Ukraine, but quite another to see video footage of it.

Fallis persuasively describes how deepfakes make it harder to form justified beliefs by "reduc[ing] the *amount of information* that videos carry to viewers."[187] "Information" is a technical term in epistemology that refers to how much some piece of evidence tells a given viewer about the world[188] and how reliably.[189] A video will typically carry less information if it is low resolution or shot from a bad angle. It also carries less information if it is less likely to portray something true. More formally speaking,

[180] Adrienne de Ruiter, *The Distinct Wrong of Deepfakes*, 34 PHIL. & TECH. 1311, 1324, 1326 (2021) (emphasis omitted) ("Non-consensual deepfakes wrong the persons they portray because they manipulate the process through which people's identity is socially constituted . . . .").

[181] Don Fallis, *The Epistemic Threat of Deepfakes*, 34 PHIL. & TECH. 623, 624–27 (2021).

[182] *See infra* Section II.B.1.

[183] *See infra* Section II.B.2.

[184] Fallis, *supra* note 181, at 624–27.

[185] *Id.*

[186] *Id.* at 624.

[187] *Id.*

[188] *See* FRED I. DRETSKE, KNOWLEDGE AND THE FLOW OF INFORMATION 3–4 (1981); Jonathan Cohen & Aaron Meskin, *On the Epistemic Value of Photographs*, 62 J. AESTHETICS & ART CRITICISM 197, 206–07 (2004).

[189] *See* BRIAN SKYRMS, SIGNALS 35 (2010) ("The key to information is moving probabilities.").

> R [some signal reflecting a piece of evidence] carries the information that S [some state of the world] when the likelihood of R being sent when S is true is greater than the likelihood of R being sent when S is false. . . . [T]he more likely it is for a signal R to be sent in the state where S is true than it is for R to be sent in the state where S is false, the more information that R carries about S.[190]

Deepfakes make it more likely that a video conveying some content about the world would exist even if that content were false: "Deepfake technology increases the probability of a false positive. . . . As a result, videos carry less information than they once did."[191] The more prevalent deepfakes become, the less information all videos—even truthful ones—will carry. This generates epistemic harm because "we cannot learn as much about the world if less information is carried by videos."[192]

Videos will carry *even less* information in courtrooms once deepest fakes are possible. The eventuality of deepest fakes will dampen information carry in two predictable ways on the Fallis framework. First, because deepest fakes will be indistinguishable from authentic media, it will become easier for them to send an undetectably false signal.[193] Second, because deepest fakes will be very easy to make, there will simply be more of them. As the ratio of deepest fakes to authentic videos—i.e., the ratio of noise to signal—increases, videos will become more likely to contain false content.[194] The courtroom context makes matters worse. Highly motivated, adversarial parties have stronger incentives to create or introduce fake content. We should expect that deepest fakes would be even more highly concentrated in courtrooms than in the wild.

Regina Rini argues that the epistemic harms of deepfakes reach far beyond digital media. Rather than start with the important epistemic role of videos, Rini begins with testimony—i.e., evidence we receive from others telling it to us: "Our collective epistemic practices are highly reliant on testimony . . . ."[195] Indeed, most of our higher order learning comes from testimony. For those of us who do not perform basic science, unearth ancient artifacts, or visit foreign countries, testimony is often the only evidence we have. I know that water is made

[190] Fallis, *supra* note 181, at 629 (emphasis omitted).

[191] *Id.* at 632.

[192] *Id.* at 633.

[193] *See id.* at 632 ("[T]he probability of a false positive depends on the viewer's ability to distinguish between genuine videos and fake videos.").

[194] *See id.* at 626 ("[D]eepfake technology threatens to drastically increase the number of realistic fake videos in circulation.").

[195] Regina Rini, *Deepfakes and the Epistemic Backstop*, PHILOSOPHER'S IMPRINT, Aug. 2020, at 1, 1.

of hydrogen and oxygen, that I have two kidneys, and that iPhones are made in China only because others have told me.

Of course, we cannot believe everything we are told. We are only justified in believing something someone tells us if we are antecedently justified in believing that person is trustworthy.[196] We might believe someone is trustworthy because we have an extended relationship with them, in the course of which they have displayed their commitment and capacity to say true things to us. For everyone else—from teachers, to book authors, to random people we ask for directions—we usually start by trusting what they say because of the social norms that govern testimony: "When a person attempts to provide testimony, she is taken to be implying that she is both *sincere* and *competent* . . . ."[197] One reason we can rely on others to follow these norms is that there are reputational consequences for misstating the truth.[198] A scholar whose articles contain falsehoods will not have many readers after the word gets out.

On Rini's account, videos play a critical role enforcing testimonial norms.[199] Her idea is that in public spaces, there is an ever-present possibility of being recorded, whether by a security closed-circuit television, a doorbell camera, or in the background of someone's TikTok montage.[200] Testimony is more trustworthy because people know there is a decent chance that what they say is being recorded and that any falsehoods could have reputational consequences. For this reason, Rini says, "Video and audio recordings function as an epistemic backstop."[201]

Rini agrees with Fallis that deepfakes make videos overall less reliable because they carry less information.[202] That itself is an epistemic harm. But, for her, the more worrisome effect is that "video and audio recordings may lose their status as acute correctors of the testimonial record."[203] Rini envisions a world in which we not only trust videos less but also each other. Deepfakes undermine the epistemic backstop of video, making the entire network of testimonial knowledge vulnerable

[196] Don Fallis, *Lying and Omissions*, *in* The Oxford Handbook of Lying 183, 191–92 (Jörg Meibauer ed., 2019).

[197] Rini, *supra* note 195, at 2.

[198] *Id.*

[199] *Id.* ("The availability of recordings undergirds the norms of testimonial practice, increasing the incentive for testifiers to speak with sincerity and competence.").

[200] *Id.* at 4 ("[W]hen we are in public urban spaces, we know that we're more likely than not covered by CCTV cameras or traipsing through the background of any number of strangers' selfie-directed phones.").

[201] *Id.* at 2 (emphasis omitted).

[202] *Id.* at 7 ("The obvious worry about deepfakes is that they will be used to propagate vivid disinformation. . . . But I think that the most important risk is . . . that increasingly savvy information consumers will come to reflexively distrust *all* recordings.").

[203] *Id.* at 8.

to collapse; "the gravest danger of deepfakes [is that] . . . [w]ithin a few years, we may have little reason to trust the testimony of strangers . . . ."[204]

It is easy to see how deepest fakes and the courtroom context would amplify Rini's concerns. As argued above, when deepest fakes are possible, videos overall will carry less information, even more so in the courtroom.[205] If Rini is right that videos serve as a critical epistemic backstop for testimony, this could be devastating for the judicial process. Trials are almost always about past events that happened out of the courtroom. This makes factfinders especially dependent on witness testimony. If witnesses become unreliable because videos can no longer credibly impeach them, it is hard to see how courts could continue to function.

2. *Deepfake Optimism*

Some philosophers are more optimistic about the resilience of our epistemic practices. Most of these philosophers place their confidence in technological interventions like those discussed above—e.g., investing in deepfake detectors or blockchain video authentication.[206] We have already shown why those approaches are unlikely to help.[207] Two epistemologists—Joshua Habgood-Coote and Keith Raymond Harris—offer more sophisticated reasons for optimism. Unfortunately, neither is up to the challenge of deepest fakes and courtrooms.

Unlike Rini, Habgood-Coote does not think video has a unique epistemic role to play vis-à-vis other sorts of media.[208] As he argues through a detailed history of manipulation in photographs, persuasive media fakery is nothing new.[209] Yet, somehow, our epistemic practices adapted so that we do still sometimes rely on photographs for forming

[204] *Id.*

[205] *See supra* notes 187–94 and accompanying text.

[206] *See, e.g.*, Luciano Floridi, *Artificial Intelligence, Deepfakes and a Future of Ectypes*, 31 PHIL. & TECH. 317, 317–18 (2018); Fallis, *supra* note 181, at 640 ("[A]nother possible strategy for increasing the amount of information that videos carry is for us to get better . . . at identifying deepfakes.").

[207] *See supra* Section I.D. Keith Raymond Harris offers additional persuasive reasons against relying on deepfake detectors, even when they are provably accurate. *See* Keith Raymond Harris, *AI or Your Lying Eyes: Some Shortcomings of Artificially Intelligent Deepfake Detectors*, PHIL. & TECH., Jan. 10, 2024, at 1, 5–6.

[208] *See* Joshua Habgood-Coote, *Deepfakes and the Epistemic Apocalypse*, SYNTHESE, Mar. 9, 2023, at 1, 3–5.

[209] *See id.* at 17 ("Forgetting the history of photographic manipulation both encourages us to think of deepfakes as a novel problem, and amplifies our perception of the seriousness of the problem."); *see also* BRITT PARIS & JOAN DONOVAN, DEEPFAKES AND CHEAP FAKES: THE MANIPULATION OF AUDIO AND VISUAL EVIDENCE 5–6 (2019) (describing deepfakes as merely "one component of a larger field of audiovisual . . . manipulation" that has "never been stable").

justified beliefs.[210] The reason, Habgood-Coote says, is that norms developed to govern photography,[211] much like the epistemic norms that Rini says govern testimony.[212] When we trust what we see in a photograph, we are not only trusting an individual photographer, but a diffuse set of epistemic practices that binds photographers.[213] Thus, "the reason why we continue to trust photograph[y] is that, in epistemic photographic practices, photo-manipulation is unprofessional, and is punished."[214]

Habgood-Coote predicts that similar norms for video creation will develop, if they are not already in effect: "I take it as given that producing inaccurate deepfakes and disseminating them as real videos is a violation of the norms of producing and disseminating videos."[215] Of course, deepfakes do exist, but the reason they do is that there are "long-running problems around the management of the norms of producing and disseminating recordings."[216] To the extent that these norms need a little encouragement from the outside, Habgood-Coote is confident that we have a good "sense of how to design better social practices."[217] He suggests removing the financial incentives for making deepfakes, particularly pornography; banning online forums where deepfakes are shared; and taking down widely used tools for making deepfakes.[218]

Setting aside the question of whether the interventions Habgood-Coote proposes are consistent with the First Amendment,[219] it is doubtful that social regulation would work for deepest fakes or courtrooms. Any ban or restriction on deepfakes will be exceedingly hard to enforce against deepest fakes, which, by definition, are indistinguishable from authentic media. Habgood-Coote presupposes there would be no "catastrophic norm flouting," but that is exactly what deepest fakes enable.[220] Even if he is right that epistemic norms around video creation

[210] Indeed, some philosophers have argued that seeing something in a photo is akin to directly perceiving it. *See, e.g.*, Dan Cavedon-Taylor, *Photographically Based Knowledge*, 10 Episteme 283, 283 (2013).

[211] Habgood-Coote, *supra* note 208, at 12 ("[T]he development of the professional identity of the documentary photographer established a practice of photography in which photographers were both trusted and trustworthy, within which manipulated photos counted as norm violations.").

[212] Rini, *supra* note 195, at 2.

[213] *See* Sandy Goldberg, *The Division of Epistemic Labor*, 8 Episteme 112, 112–13 (2011); Habgood-Coote, *supra* note 208, at 7 ("[W]e rely on a set of information-dissemination practices . . . .").

[214] Dominic McIver Lopes, Four Arts of Photography 110 (2016).

[215] Habgood-Coote, *supra* note 208, at 7.

[216] *Id.* at 18 ("[W]hether [deepfake] videos are created . . . is a matter of the social context in which this technology is deployed.").

[217] *Id.*

[218] *Id.*

[219] *See supra* Section I.A.

[220] Habgood-Coote, *supra* note 208, at 8.

will develop, like those that govern professional photographers, there is no reason to think those norms would extend to the courtroom. Most plaintiffs and defendants are not professional videographers, so they would have no reason to know or follow the relevant norms. Even if they did, the adversarial context can provide very strong incentives for flouting norms when there is little chance of detection.

Keith Raymond Harris has different reasons for concluding that "concerns that deepfakes will bring about epistemic catastrophe are overblown."[221] Harris's important insight is that "the evidential power of video derives not solely from its content, but also from its source."[222] A video handed to us from a trusted source holds a different epistemic value than a video that comes from an unknown or untrusted source.[223] Ordinary people can avoid broader digital media skepticism "by taking a skeptical attitude [only] toward video footage that does not come from trusted sources . . . [while] continu[ing] to rely upon video footage from trusted sources."[224]

Although Harris's approach to deepfakes strikes us as a step in the right direction, it does not have internal resources for handling deepest fakes. It is clear from Harris's framework that the "source" of a video is the person or entity who provides or "present[s]" the video, not the person who records the video.[225] For example, if a news station plays footage provided from an informant, the source for the viewers is the news station rather than the informant. As Harris anticipates, deepest-fake technology could "generate fabricated video footage [falsely] depicting [that it comes from] a trusted [source]."[226] For example, a TikTok video could falsely depict a CNN anchor introducing a fake video. Harris offers only a partial solution. Ordinary people may learn to access videos directly from the channels the videos purport to come from—e.g., by tuning into CNN rather than watching TikTok.[227] However, this "is of no utility to sources themselves."[228] Although channels

[221] Keith Raymond Harris, *Video on Demand: What Deepfakes Do and How They Harm*, 199 Synthese 13,373, 13,374 (2021).

[222] *Id.*

[223] *Id.* Fallis makes a related point, though he does not expand much on it. *See* Fallis, *supra* note 181, at 640 ("[E]ven without laws against deepfakes, the evening news is subject to normative constraints. Thus, we can try to identify those videos that still carry a lot of information.").

[224] Harris, *supra* note 221, at 13,380.

[225] *See id.*

[226] *Id.* at 13,383; Fallis, *supra* note 181, at 640 ("Purveyors of deepfakes can try to make it difficult for people to determine whether a video comes from a source that is subject to normative constraints.").

[227] Harris, *supra* note 221, at 13,384 ("The present concern draws attention to the oft-neglected significance of what we might call *channels* of information. . . . While purveyors of deepfakes might exploit patterns of trust by using certain likenesses and logos, they cannot easily inject deepfakes into particular channels.").

[228] *Id.* at 13,382.

"can continue to rely on their own video footage"[229]—e.g., CNN can rely on footage its own reporters take—they must reckon with the threat of digital media skepticism for everything else.

It is also difficult to see how Harris's approach would translate to the courtroom. He says little of how we come to trust a source or channel. Presumably, trust is the sort of thing that builds over time through repeated interaction. In the adversarial courtroom setting, trust is usually in short supply. There are no court-sanctioned media sources, let alone media channels. So, although Harris is certainly right that that a video's evidential value depends in part on the video's source, he does not offer tools for warding off digital media skepticism in the courtroom.

C. *Philosophical Takeaways*

By now, the full threat that deepfakes pose should be clear. Digital media have become an indispensable part of our epistemic practices. Because each of us can only directly learn about a very narrow slice of reality, we must learn everything else indirectly from other sources. Digital media are one such source because they can provide a durable, accurate record of events that took place at distant places and times. Deepfakes destabilize this pathway to knowledge by pushing us toward *digital media skepticism*. Once we learn that digital media content can be persuasively faked, we may "come to reflexively distrust *all* recordings."[230] If we become digital media skeptics, we will not be able to know nearly as many things as we previously could.

To make matters worse, digital media skepticism might not be an overreaction. Prominent views in epistemology explain why, under the right conditions, digital media skepticism might become a rational response. For purposes of this Article, we have assumed that two such conditions hold. First, we suppose that deepest fakes—which are indistinguishable from genuine media—will eventually be possible, plentiful, and costless to make. Second, we focus our discussion on the adversarial courtroom, where interpersonal trust is extremely low. Under these conditions, there are strong evidentialist and reliabilist arguments for digital media skepticism. *Evidentialists* believe that we come to know things by forming justified beliefs based on sufficient evidence.[231] Although digital media might once have been good evidence for believing the events they portray, the existence of deepest fakes *defeats* digital media's justificatory power. As deepest fakes become increasingly prevalent, the *information* digital media carry, or the *signal* they send,

[229] *Id.* at 13,383.

[230] Rini, *supra* note 195, at 7.

[231] *See* Deborah M. Hussey Freeland, *Speaking Science to Law*, 25 GEO. INT'L ENV'T L. REV. 289, 299 (2013).

weakens because the information environment becomes polluted with indistinguishable *noise*—i.e., false signals or misleading content. Eventually, the truth signal that digital media send will become so low that the rational response will be to distrust just about all digital media. *Reliabilists* argue that we come to know things by forming justified beliefs using processes that tend to generate true beliefs.[232] When deepest fakes are possible and interpersonal trust is minimal, forming beliefs by viewing digital media becomes a highly unreliable process. In such contexts, reliabilism would also recommend digital media skepticism.

This marks the turning point of the Article. Digital media skepticism seems all but inevitable from both philosophical and legal perspectives. Purported solutions melt away in the face of deepest fakes and the adversarial context of the courtroom. Yet, in what follows, we argue that evidence law has long had resources for staving off digital media skepticism and that evidence law holds lessons for responsible media consumption in ordinary life.

We depart from prior philosophical work on deepfakes in a critical respect: We center testimony. *Testimony* is the type of evidence we gain from other people when they tell us verbally or in writing that some state of the world obtains.[233] As mentioned above, both evidentialists and reliabilists think testimony is important. "So much of what we know about the world, e.g., history, science, politics, one another, etc., comes from the testimony of others."[234] When your science teacher told you that the earth is round, you probably formed a justified belief about the shape of the planet. You did not have to see the horizon's curvature yourself or view photos taken from space. You acquired good evidence and employed reliable belief-forming processes.

Philosophers writing about deepfakes generally ignore testimony, or they bring it up only to diminish it. Recall that Rini believes our everyday epistemic practice of relying on others' testimony only works because video recordings can help us detect people who lie.[235] She envisions a world of cascading skepticism in which deepfakes lead to digital media skepticism and digital media skepticism leads to testimony skepticism: "[R]ecordings will be demoted . . . to sources of mere testimonial evidence. And if they are simply just another source of testimony, then they cannot be relied upon to correct or regulate testimonial practice."[236]

[232] *See id.*

[233] *See* Nick Leonard, *Epistemological Problems of Testimony*, *in* Stanford Encyclopedia of Philosophy § 7 (Edward N. Zalta & Uri Nodelman eds., spring 2023 ed. 2023), https://plato.stanford.edu/archives/spr2023/entries/testimony-episprob/ [https://perma.cc/5K9Y-JBWY].

[234] *Id.*

[235] Rini, *supra* note 195, at 8 ("Within a few years, we may have little reason to trust the testimony of strangers, as the norms securing their anticipated cooperation come gradually undone.").

[236] *Id.* at 10 (emphasis omitted); Habgood-Coote, *supra* note 208, at 6 ("Once we become aware of the possibility of deepfakes, when we form beliefs based on videos, we must either

We see things differently. Like most epistemologists,[237] we view even "mere" testimony to be an indispensable source of evidence. Testimonial evidence existed before digital media, and it will survive the advent of deepest fakes.

Indeed, we argue that tethering digital media closer to testimony is the best way to avoid digital media skepticism. The reliability of human testimony can save deepfakes, rather than vice versa. This flips Rini's justificatory picture on its head. The misstep we see in the philosophy of deepfakes is that epistemologists often compare deepfake technology to Photoshop.[238] They puzzle over how humanity managed to avoid photo skepticism in the face of highly persuasive photo-manipulation techniques. We frame our discussion in terms of a much more ancient form of deceit, humanity's original fakery: the simple fib. Ever since humans could represent states of the world through language, they have also used their words to *mis*represent states of the world. There are diverse types of misrepresentation, and a speaker's intention plays an important role in distinguishing between them:[239] from lies that are designed to deceive, to creative expressions that are designed to entertain or educate, to accidental inaccuracies that arise because of speaker incompetence. Humans have developed epistemic tools for sorting good from bad testimony. In what follows, we contend that these are the key for assessing the impact deepfakes will have.

III. Evidence Law of Fakeries and Other Creative Content

As the guiding framework for nearly every inquiry into accuracy, authenticity, relevance, and reliability in the structured court environment, the rules and history of evidence hold clues for how people can adapt to a real-world media environment saturated with deepfakes. Optimistic evidence scholars seem to think that if they can just find the right rules, a judge mechanically applying them will filter out deepfakes and expose jurors only to media that provide a direct window into truth. Pessimists agree with the goal but despair of ever finding such rules. They worry that the looming possibility of deepfakes will drive jurors

extend our trust to the videographer, making videographic knowledge akin to knowledge from testimony, or rely on background beliefs about the likelihood of faking, making it into a kind of inferential knowledge. Either way, videographic knowledge loses its distinctive character as non-interpersonal knowledge . . . ." (footnote omitted)).

[237] *See, e.g.*, Rini, *supra* note 195, at 2 ("I seem to have a default justification to accept testimony as evidence . . . .").

[238] *Id.* at 12; Habgood-Coote, *supra* note 208, at 12.

[239] *See* Ruiter, *supra* note 180, at 1313 ("The moral evaluation of specific deepfakes depends on . . . the intent with which the deepfake was created.").

toward digital media skepticism and that courts' factfinding mission will lose all credibility or meaning.

We think the optimistic position is naive and the pessimistic position is shortsighted. Both overlook the essential and powerful intermediating role of human judgment. Optimists' quest for rule-based guarantees will fail because the deepest fakes will thwart any generic truth filter. Pessimists' predictions will fail because human judgment stands as a bulwark between distrust and skepticism.

The developments in evidentiary practice that we envision are contemporary retellings of a familiar story. Since the beginning of modern evidence practice—marked by a shift from relying on divine omniscience to relying on human judgment[240]—the law has grappled with the challenge of handling untrue and deceptive evidence. Fibs and forgeries of all sorts are a potent concern when litigants have life, limb, and purse on the line. But excluding every category of newly manipulable evidence would leave courts with precious little to consider. Instead, the law has continuously striven to include all but the riskiest evidence in an adversarial process that culminates in jurors' commonsense assessments of credibility and probative value. Law's future answers to the threat of deepfakes has been foretold in the history of its response to two key evidentiary challenges: lying witnesses[241] and mechanically recorded evidence.[242]

A. *The History of Lies and Deception*

Deception through deepfakes is the novel expression of an exceedingly old legal challenge. For as long as disputes have been resolved through litigation, courts have needed a method for distinguishing truthful testimony from lies. The precursor to most trials is, after all, conflicting claims about who did what. The plaintiff says that the defendant stole a horse; the defendant denies it. If both sides swear on their oath that their respective statement of the facts is true, then the court is handed the unenviable task of deciding which of two earnestly pledged factual statements is true—or, to put it another way, to decide which side is lying.[243]

Everyone who is capable of providing testimony is capable of lying, and the trial context can supply strong motivation to do so. Who would not claim innocence if it bought them a slim chance of escaping capital punishment? In a civil context, what private party does not feel at least the urge to edit and exaggerate their story to the benefit of their litigation

[240] *See infra* Section III.A.

[241] *See infra* Section III.A.

[242] *See infra* Section III.B.

[243] This is particularly apparent in the context of a criminal defendant who vigorously claims their innocence. There is no logical separation between the conclusion that such a defendant is guilty as charged and that the defendant is also guilty of perjury.

posture? Even a third-party witness may feel compelled to shade or embellish their testimony in order to protect a party—or themself—from its consequences. These propositions sound prosaic to the modern ear, but they reflect a surprisingly fundamental and enduring challenge for our legal system.

If every witness could be lying, and if many have strong motivations to do so, then what justification can there be for favoring one witness's story over another's? Grave consequences turn upon the answer. Yet few of us are born with any innate ability to separate truth from fiction when handed equally compelling but conflicting accounts about what happened.[244] The intertwined histories of trials and the law of evidence are substantially a history of trying to reach just results despite the limits of human lie detection.

In the eleventh-century precursors to modern trials, for example, the need for mortal judgement between conflicting statements was often relieved by performances tinged with interpretation as divine judgment.[245] Let us suppose that a member of a medieval community had been credibly accused of stealing another's horse. Perhaps motivated by a desire to escape punishment—which might be as severe as death or mutilation—the accused emphatically swears their innocence. To avoid the stalemate that would result from opposing sworn statements by the accused and the accuser, courts would task the accused with establishing their innocence through something like trial by the ordeal of hot iron.[246]

In this ordeal, a judge or priest would order the accused to pick up and carry a searing iron weight for a length of nine paces.[247] After this performance, the accused's hands would be bandaged for three days.[248] Once the waiting period was up, the court would reassemble to unwrap and inspect the hands.[249] Infected, pustulant burns were proof of a guilty soul and thus established that the accused had falsely claimed their innocence.[250] Healed—or better yet, undamaged—hands proved the purity of the accused's soul and thus their innocence.[251]

[244] *See infra* notes 277–79 (discussing the dismal empirical record on human skill in lie detection).

[245] *See generally* ROBERT BARTLETT, TRIAL BY FIRE AND WATER (1986) (examining the development and decline of trial by ordeal from barbarian law codes through modern Europe and America).

[246] This example is meant only as an illustration. Specific details on when and what ordeals would be ordered seem to have varied over time and by jurisdiction. *See id.* at 2.

[247] Margaret H. Kerr, Richard D. Forsyth & Michael J. Plyley, *Cold Water and Hot Iron: Trial by Ordeal in England*, 22 J. INTERDISC. HIST. 573, 588 (1992).

[248] *Id.*

[249] BARTLETT, *supra* note 245, at 1.

[250] *Id.*

[251] *Id.*

This ordeal parallels other factfinding devices of the time. In the ordeal of cold water, a person's honesty was established when, bound and lowered into water, they did not float but were received by the water and sank.[252] In a trial by judicial combat, the party whose account was truthful would be propelled by their purity to triumph in battle against their opponent.[253] In every case, the conclusion of who was telling the truth was entrusted to divinity and to the spirit.[254] God revealed which side was telling the truth, obviating the need for mortal judgment,[255] which rested the order to dispatch legal sanction upon the highest and most unquestionable authority.

Arresting as this period of trial practice was in legal history, it came to an end in 1215 when the church withdrew its endorsement of trial by ordeal.[256] The historic record is regrettably sparse on what exactly trial practice looked like during the next several hundred years, but a few details are reasonably certain. Without the aid of divine intervention, juries were forced to step forward into something analogous to their current role.[257] The job of finding legal truth thus became a human undertaking. Still, a variety of practices continued to relieve juries from needing to process directly conflicting testimonial narratives.

One such practice was an apparently strong preference for documentary evidence during this period.[258] Documents, especially those sealed and trustworthy by virtue of how they had been created, were preferred and emphasized over live witness testimony, especially in contract disputes.[259] Sir Geoffrey Gilbert's respected treatise on evidence,

[252] *See* Kerr et al., *supra* note 247, at 582–83.

[253] *See* BARTLETT, *supra* note 245, at 103–26. *See generally* GEORGE NEILSON, TRIAL BY COMBAT (The Lawbook Exch., Ltd. 2000) (1890) (recounting the historical belief that divine judgment guided the outcome of combat, with victory awarded to the juster cause).

[254] Paul R. Hyams, *Trial by Ordeal: The Key to Proof in the Early Common Law* ("Unilateral ordeals, oaths, and duels share one important factor. All three methods of proof purport to work by revealing God's judgment."), *in* ON THE LAWS AND CUSTOMS OF ENGLAND 90, 92 (Morris S. Arnold ed., 1981); *see* NEILSON, *supra* note 253, at 111 ("In such circumstances the accused was bound to purge himself by the judgment of God, viz., by the hot iron if a freeman, by water if a villein, according to the divers conditions of men.").

[255] Hyams, *supra* note 254, at 92–93.

[256] *See* Roger D. Groot, *The Early-Thirteenth-Century Criminal Jury* ("The most important event in the history of the criminal jury was the abolition of the ordeal by edict of the Roman church in 1215."), *in* TWELVE GOOD MEN AND TRUE 3, 3 (J.S. Cockburn & Thomas A. Green eds., 1988); *see also* BARTLETT, *supra* note 245, at 34 (commenting that the ordeals were "everywhere vestigial" by 1300).

[257] *See* George Fisher, *The Jury's Rise as Lie Detector*, 107 YALE L.J. 575, 585–86 (1997) ("The occasion of [the] sudden birth of trial by jury was the sudden death of trial by ordeal.").

[258] *See* John H. Langbein, *Historical Foundations of the Law of Evidence: A View from the Ryder Sources*, 96 COLUM. L. REV. 1168, 1181 (1996) ("The law of evidence in its infancy was concerned almost entirely with rules about the authenticity and the sufficiency of writings.").

[259] *Id.* at 1183 ("The preference for written evidence extended back to the Middle Ages, and was particularly apparent in contract and conveyancing. The judges determined by the fourteenth

published posthumously in 1754, emphasized the identification, authentication, and epistemic ranking of documentary evidence in detail, while devoting comparatively little attention to the subject of witness testimony.[260] Legal historians report this treatment to be consistent with the surviving record of trial practice at this time.[261]

Even more important were a variety of rules that simply prohibited all testimony from witnesses deemed likely to lie under oath. "Incompetent" witnesses in this regime included criminal defendants,[262] both parties to civil disputes,[263] the spouses of parties,[264] others with personal interest in the outcome,[265] children,[266] convicted criminals,[267] and atheists.[268] Though the details of the rationale varied from one category to the next, the basic reasoning was always the same: Because the oath of an incompetent witness could not be trusted to compel truthful testimony, the witness was prophylactically stricken from the stand.[269] This was an act of generosity to the jurors (who would not be challenged with conflicting testimony)[270] as well as to the witness (who would not be tempted into tainting their soul with perjury).[271]

century that only contracts written and sealed would be actionable under the writ of covenant. . . . The legal system that endured into [the 1750s] had exhibited a centuries-long proclivity for suppressing resort to oral evidence at jury trial in civil matters." (footnote omitted)).

[260] T.P. Gallanis, *The Rise of Modern Evidence Law*, 84 Iowa L. Rev. 499, 506–07, 506 n.37 (1999) (noting the comparative emphasis on written over unwritten evidence in Gilbert's treatise).

[261] *See* Langbein, *supra* note 258, at 1183 ("Ryder's trial practice reflects the preoccupation with written evidence that we find in Gilbert and the other eighteenth-century writers." (footnote omitted)); Gallanis, *supra* note 260, at 511 ("Evidentiary practice in civil trials focused principally on questions of written evidence.").

[262] *See* Fisher, *supra* note 257, at 624.

[263] *See* Langbein, *supra* note 258, at 1184–86.

[264] Fisher, *supra* note 257, at 624.

[265] G.S., *Competency of Witnesses*, 10 Am. L. Reg. 257, 265 (1862) ("The rule is that a present interest in the event of a suit excludes the witness. But it must be a *certain* interest, and then no matter how small it is.").

[266] Gallanis, *supra* note 260, at 507 (paraphrasing Gilbert's description of an incompetence category for "those lacking in discernment," which included children under a certain age as well as those deemed to have intellectual disabilities).

[267] G.S., *supra* note 265, at 264 (summarizing the rule that "judgment against any person for treason, felony, or the *crimen falsi*, renders him incompetent to testify").

[268] Fisher, *supra* note 257, at 624.

[269] *See id.* at 625 (describing these rules as "declaring certain witnesses to be likely liars as a matter of law").

[270] *See id.* at 626 ("[C]ompetency rules did the work of lie detecting, so the jury did not have to.").

[271] Gilbert's treatise puts the matter in essentially these same terms:

> Now where a man who is interested in the matter in question, would also prove it, it is rather a ground for distrust than any just cause of belief; for men are generally so short sighted, as to look at their own private benefit which is near to them, rather than to the good of the world, that is more remote; therefore, from the nature of human passions and actions, there is more reason to distrust such a biassed testimony than to believe it;

Finally, a variety of additional rules and instructions apparently stood ready to relieve jurors of the need to identify a lie in cases in which conflicting sworn testimony did manage to come before the court. One example was an instruction sometimes given to jurors that they should attempt, where possible, to reconcile sworn testimony so that their interpretation did not require assuming that either witness was lying.[272] Another was an occasional suggestion that jurors should resolve conflicts in sworn testimony by counting the number of witnesses for and against a proposition rather than by trying to evaluate the individual credibility of each witness.[273]

Like trial by ordeal before it, this approach to trial process eventually came to an end. Starting in the 1840s, a series of legislative acts on both sides of the Atlantic toppled one competency rule after another.[274] Concerns about fairness, the need for information, and other less obvious considerations motivated these retractions.[275] But these concerns could not plausibly have escaped legal thinkers during the long tenure of the competency rules. A better explanation of the previous practice seems to be a kind of testimonial skepticism—i.e., the simple doubt that a jury of laypeople would be able to distinguish truth from well-presented fiction in the trial context. If the jury's handling of contradictory testimony was really nothing more than a random guess about who was telling the truth, then plausibly the accuracy of verdicts in the presence of conflicting testimony was no better than the quality of verdicts in its absence.

Contemporary trial practice evinces a willingness to entrust judges and juries with the testimony of every witness who appears. This reflects recent confidence in the power of human agents to sort truth from fiction. H. Richard Uviller once said, "At the heart of our adversarial mode of adjudication lies the *assumption* that trial jurors—a fair mix of ordinary, relatively openminded folk—can from across the jury rail

> it is also easy for persons who are prejudiced and prepossessed, to put false and unequal glosses for what they give in evidence, and therefore the law removes them from testimony, to prevent their sliding into perjury; and it can be no injury to truth, to remove those from the jury, whose testimony may hurt themselves, and can never induce any rational belief.

Geoffrey Gilbert, The Law of Evidence 106 (Garland Publ'g 1979) (1754).

[272] *See* Fisher, *supra* note 257, at 624–33 (describing "the Rule of *Bethel's Case*").

[273] *Id.* at 653 ("Almost every major treatise suggested that whenever jurors faced the task of choosing between conflicting oaths, they should tend to give more credit to the side that produced the greater number of witnesses.").

[274] *See id.* at 658–59 (listing several of the relevant acts); G.S., *supra* note 265, at 257 (observing at the time of writing that "[p]ractically now in the English courts all persons are competent witnesses, their credibility being left to the jury").

[275] *See* Fisher, *supra* note 257, at 659–71; *see also id.* at 671–97 (describing how racial considerations interacted with rules of witness competency).

distinguish liars from truthtellers."[276] Yet, decades of research show fairly consistently that most people do only a little better than a coin flip in discerning truthful statements from lies.[277] This is true even when factoring in demeanor and other nonverbal evidence supposedly indicative of a witness's state of mind.[278] Indeed, some research suggests that a focus on demeanor cues does more harm than good in helping jurors identify the truth.[279] It seems jurors are no better at detecting lies than they are at detecting sophisticated deepfakes.

How, then, could we possibly justify courts' enduring reliance on witness testimony? The epistemological frameworks discussed in Part II can help diagnose the problem. To perform their task well, jurors need to form justified beliefs about the content of witness testimony. Although a witness's words and demeanor may be important evidence, studies show that they are insufficient.[280] The fact that some witnesses are skilled liars defeats the evidentiary value of witness words and demeanor by reducing the signal that they can send. On a systems level, this puts court verdicts at risk of being epistemically unjustified because they result from an unreliable process reliant on jurors assessing witness testimony.

The solution is to make the courts' process more reliable by putting jurors in a better position to form justified beliefs about witness testimony. Ultimately, that means providing jurors more and better evidence on what witnesses say. One possibility—analogous to some proposals

[276] H. Richard Uviller, *Credence, Character, and the Rules of Evidence: Seeing Through the Liar's Tale*, 42 Duke L.J. 776, 780 (1993) (emphasis added); *see also id.* at 776–77 ("Our faith in the adversary system . . . depends in large measure on our confidence that, assisted by courtroom procedure, our jurors will usually return a verdict consistent with the historical fact.").

[277] *See* David M. Markowitz, *Self and Other-Perceived Deception Detection Abilities Are Highly Correlated but Unassociated with Objective Detection Ability: Examining the Detection Consensus Effect*, Sci. Reps., July 30, 2024, at 1, 2 ("Overwhelming evidence in the deception detection literature suggests that on average, people are often slightly greater than chance at lie-truth judgments. Deception detection accuracy tends to hover around 54%, with truths being evaluated more accurately than lies because people are truth-biased." (footnote omitted)); Charles F. Bond, Jr. & Bella M. DePaulo, *Accuracy of Deception Judgments*, 10 Personality & Soc. Psych. Rev. 214, 217 (2006) (describing a large meta-analysis of deception studies in which average truth detection was measured at slightly above fifty percent).

[278] *See* Aldert Vrij & Jeannine Turgeon, *Evaluating Credibility of Witnesses—Are We Instructing Jurors on Invalid Factors?*, 11 J. Tort L. 231, 233–37 (2018).

[279] *See id.* at 237 (noting little evidence to support the "myth about the strong relationship between nonverbal behavior and deception"); Danielle Andrewartha, *Lie Detection in Litigation: Science or Prejudice?*, 15 Psychiatry, Psych. & L. 88, 92–93 (2008) (noting that nonverbal behavior like apparent nervousness is an especially questionable indicator of deception in the unnatural and confrontational setting of courtroom testimony); Olin Guy Wellborn III, *Demeanor*, 76 Corn. L. Rev. 1075, 1075, 1091–94 (1991) (concluding that ordinary observers often cannot effectively use demeanor to assess truthfulness and that overreliance on demeanor can be misguided).

[280] *See* Vrij & Turgeon, *supra* note 278, at 235–37; Wellborn III, *supra* note 279, at 1075.

regarding deepfakes[281]—would be to recruit expert lie detectors. Fortunately, courts did not go that route because experts do not perform much better than laypeople.[282] Rather, courts realized that there is evidence beyond a witness's words and demeanor that jurors are equipped to evaluate. Testimony does not exist as an isolated datapoint, sealed off from the justificatory web of evidentiary interdependencies that connects all truth. It is situated within a context, and threads from that context can provide additional evidence. Some threads will reveal consistencies and inconsistencies internal to the testimony. Others show whether the testimony is consistent or inconsistent with truths external to the testimony itself. All that is needed is a process that reliably generates such evidence and puts it before the jurors. That process is the adversarial trial.[283]

Within the adversarial trial, opposing council spin for the jury the web of evidentiary interdependencies by presenting evidence, eliciting testimony, and—importantly—testing witness testimony by cross-examination. Jeremy Bentham once proclaimed, "Against erroneous or mendacious testimony, the grand security is cross-examination: cross-examination, by which, if the individual facts charged are false, true ones . . . may be brought out against them."[284] Wigmore opined that "no safeguard for testing the value of human statements is comparable to that furnished by cross-examination, . . . [which] is beyond any doubt the greatest legal engine ever invented for the discovery of truth."[285] Of more recent vintage, the U.S. Supreme Court has described cross-examination as "the principal means by which the believability of a witness and the truth of his testimony are tested."[286]

[281] *See supra* notes 88–89 and accompanying text.

[282] Paul Ekman & Maureen O'Sullivan, *Who Can Catch a Liar?*, 46 AM. PSYCH. 913, 913 (1991) (summarizing twenty years of literature as providing little reason to trust human lie-detection capabilities and reporting a study in which even professionals whose job involved lie detection typically fared little better than chance).

[283] *See generally* Ronald J. Allen, Michael S. Pardo, William J. Lawrence & Christopher K. Smiciklas, *Minimal Rationality and the Law of Evidence*, 115 J. CRIM. L. & CRIMINOLOGY 269 (2025) (arguing that much of modern evidence law supplies only foundational standards and safeguards, with the role and incentives of parties in an adversarial process supplying the principal guarantee that strong evidence and arguments will ultimately be produced).

[284] 5 JEREMY BENTHAM, RATIONALE OF JUDICIAL EVIDENCE 212 (Garland Publ'g 1978) (1827); *see also* 2 JEREMY BENTHAM, RATIONALE OF JUDICIAL EVIDENCE 230 (Garland Publ'g 1978) (1827) ("Mendacious invention, then, having been either prevented, or encompassed with dangers, by the *vivâ voce* questions followed immediately by the *vivâ voce* answers . . . ."); 2 BENTHAM, *supra*, at 231 (arguing that rapid cross-examination, with consequently little time for careful fabrication, is "the only remedy" for mendacious invention).

[285] 2 JOHN HENRY WIGMORE, A TREATISE ON THE SYSTEM OF EVIDENCE IN TRIALS AT COMMON LAW § 1367, at 1697 (1904).

[286] Davis v. Alaska, 415 U.S. 308, 316 (1974).

Confidence in this power of cross-examination is justified, to some extent, by the robust rules of evidence that have developed since Gilbert's time to limit and focus witness testimony. The strictures of relevancy[287] and prohibition on character reasoning[288] discourage testimony from wandering away from material facts. Limits on the introduction of hearsay evidence force litigants to put witnesses before the jury.[289] A formidable machinery of impeachment and rehabilitation stands ready to test the credibility of every witness who takes the stand.[290] Layered atop these testimonial screens, the threat and practice of skillful cross-examination surely does provide courts and juries with some impressive tools for spotting attempted deception.

The lesson is not that cross-examination is an unfailing engine of truth. It is not. Deepfakes are lies, and lies are efforts at deception. Like every lie that has ever been told, the production of a deepfake reveals at least the liar's belief that there is a plausible chance that others will be fooled. But in evaluating the evidentiary challenge posed by deepfakes, we cannot lose sight of the evidentiary challenges posed by *all* deceptive testimony. The possibility that every statement could be a lie is a challenge that has endured the centuries, little diminished by anything the legal system has thought to throw at it. Rather than succumb to skepticism or minimize the lay juror's role, courts have developed adversarial procedures that tend to generate contextual information for enhancing jurors' truth-finding function.[291]

B. The History of Photographs and Recordings

A series of such procedural developments with particular relevance to deepfakes arose during courts' early struggle to manage the presentation of photographs and other mechanical recordings of reality as evidence. That struggle dates, unsurprisingly, to the proliferation of these technologies between the 1850s and 1950s. In the domain of recorded images, photography's reliance on comparatively convenient and affordable media soon won out over the daguerreotype's coated copper plates.[292] By the 1870s, it appears that nearly everyone, from all

[287] *See* Fed. R. Evid. 401–402 (defining relevance and prohibiting irrelevant evidence).

[288] *See* Fed. R. Evid. 404 (prohibiting character evidence to prove conduct, with exceptions for specific purposes).

[289] Obviously, this forcing function is tempered by the many exemptions from, and exceptions to, the hearsay rule. *See* Fed. R. Evid. 801–807.

[290] *See* Fed. R. Evid. 607–609, 613 (governing witness impeachment, including character for truthfulness and prior inconsistent statements).

[291] *See supra* note 283.

[292] *The Daguerreotype Medium*, Libr. Cong., https://www.loc.gov/collections/daguerreotypes/articles-and-essays/the-daguerreotype-medium [https://perma.cc/26AY-E7ZS] (last visited Oct. 21, 2025).

walks of life, had either sat for a photograph or at least seen photographs that had been taken of friends, family, and familiar places.[293] Audio recording devices developed over a similar timeline.[294] For technical and practical reasons, however, they found fewer applications in trial evidence before the proliferation of portable magnetic-tape recording devices in the 1950s.[295] Video recording matured and expanded in usage around the same time.[296] For our purposes, it is enough to trace the history of photographs as evidence, because the themes and principles generalize to the other types of recordings.[297]

The reason that photographs posed—and still pose—a challenge for courts and the law of evidence is that they fit imperfectly into a trial process fortified against the challenges of written documents and live witness testimony. Familiar infirmities in human expression—errors in perception, memory, mendacity, and narration—were all arguably addressed by the developing law of evidence at the time photographs arrived on the scene.[298] But this new technology deviated just enough from human expression that it presented uncomfortable problems. Like all human expression, recordings could be manipulated to deceive. And in all but the rarest cases, photographs were intractably tethered to witness testimony;[299] at a minimum, a witness needed to supply the context in which the photograph would be understood and interpreted.[300] But unlike human expression, photographs held claim to special capacities like mechanical objectivity, superhuman perceptive accuracy, and

[293] *See* Udderzook v. Commonwealth, 76 Pa. 340, 353 (1874) ("The Daguerrean process was first given to the world in 1839. It was soon followed by photography, of which we have had nearly a generation's experience. It has become a customary and a common mode of taking and preserving views as well as the likenesses of persons, and has obtained universal assent to the correctness of its delineations.").

[294] *The Origins of Sound Recording*, Nat'l Park Serv. (Mar. 29, 2023), https://www.nps.gov/edis/learn/historyculture/origins-of-sound-recording.htm [https://perma.cc/D6BB-KYAR].

[295] *See* Robert C. Maher, Principles of Forensic Audio Analysis 29 (2018) ("The first portable recorders using magnetic tape appeared in the 1950s, and soon these devices were used to obtain clandestine recordings of interviews and wiretaps, as well as to record interrogations and confessions.").

[296] Judith Keilbach, *Instant TV: The Forgotten History of Video Tape Recording (and the Coverage of the Eichmann Trial)*, TMG J. Media Hist., June 26, 2024, at 1, 1–2.

[297] *See, e.g.*, Maher, *supra* note 295, at 67 (summarizing the legal treatment of audio recordings); Bureau of Just. Assistance, U.S. Dep't of Just., Video Evidence: A Primer for Prosecutors (2016), https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/final-video-evidence-primer-for-prosecutors.pdf [https://perma.cc/W6GJ-JTL2] (summarizing the legal treatment of video recordings).

[298] *See, e.g.*, Edmund M. Morgan, *Hearsay Dangers and the Application of the Hearsay Concept*, 62 Harv. L. Rev. 177, 177–78 (1948) (describing these infirmities in the context of hearsay evidence); Laurence H. Tribe, *Triangulating Hearsay*, 87 Harv. L. Rev. 957, 958–61 (1974) (same).

[299] *See, e.g.*, J.A.J., *The Legal Relations of Photographs*, 8 Am. L. Reg. 1, 5 (1869).

[300] *See* Jennifer L. Mnookin, *The Image of Truth: Photographic Evidence and the Power of Analogy*, 10 Yale J.L. & Humans. 1, 9 (1998).

near-perfect recall.[301] These features made them obviously and alarmingly potent evidence.[302]

Themes of the latter sort—scientific objectivity, precision, and unbiased truth telling—were common in much of the early discussion of photographs as evidence. Writing in 1869, one author effusively propounded the value of photographs specifically for their mechanical advantages over live witness testimony: "The photographic apparatus never intentionally falsifies nor do its products ever so fade as to distort the image they present, as do the figures of things committed to the treacherous memory of men."[303] Similar statements can be found in other commentary of the time.[304] Jennifer Mnookin summarizes this then-prevalent perspective succinctly: "[T]he photograph was not merely evidence, but the best kind of evidence imaginable: mechanical, automatic, and not subject to those biases and foibles that may cloud human judgment."[305]

Sympathetic judges saw no obstacle to admitting photographs as evidence. In 1882, for example, the Supreme Court of Georgia easily brushed aside a defendant's objections when a photograph of a victim's cut throat was introduced as evidence:

> [T]he character of the wound was important to elucidate the issue; the man was killed and buried, and a description of the cut by witnesses must have been resorted to; *we cannot conceive of a more impartial and truthful witness than the sun*, as its light stamps and seals the similitude of the wound on the

[301] *Id.* at 2, 18.

[302] *See id.* at 4 ("In the second half of the nineteenth century, two competing paradigms governed the understanding of the photograph. One emphasized its ability to transcribe nature directly, while the other highlighted the ways in which it was a human representation. From the first perspective, the photograph was viewed as an especially privileged kind of evidence; from the second perspective, the photograph was seen as a potentially misleading form of proof.").

[303] *See* J.A.J., *supra* note 299, at 5, 6–7 ("[I]f a difference exist, [sic] should we not give the greater credence to the photograph, whose testimony, we know, is perfectly truthful and generally commensurate with the fact, while that of the vouching witness, and also of the witness called to speak to the question of identity, may be mistaken or perjured?"); *see also* Mnookin, *supra* note 300, at 2 ("Seeing a photograph almost functions as a substitute for seeing the real thing.").

[304] *See, e.g.*, Rodney G.S. Carter, *"Ocular Proof": Photographs as Legal Evidence*, 69 ARCHIVARIA 23, 27 (2010) ("From the mid-nineteenth century, and continuing well into the latter part of the twentieth century, a dominant strain of the discourse surrounding photography centred on its ability to objectively reproduce what was before the lens. Given its technological origins in optics and chemistry, photography was viewed as being the product of a scientific, and therefore truthful, process, and the earliest texts announcing the invention of photography in France and Britain emphasize its mechanical nature." (footnote omitted)); Mnookin, *supra* note 300, at 17 ("In the inaugural volume of the *Philadelphia Photographer*, one author described how the camera 'sees everything and it represents just what it sees. It has an eye that cannot be deceived and a fidelity that cannot be corrupted.'" (quoting H.J. Morton, *Photography as an Authority*, 1 PHILA. PHOTOGRAPHER 180, 181 (1864))).

[305] Mnookin, *supra* note 300, at 19.

> photograph put before the jury; it would be more accurate than the memory of witnesses, and as the object of all evidence is to show the truth, *why should not this dumb witness show it*? Usually the photograph is introduced to prove identity of person, but why not to show the character of the wound? In either case it is evidence; it throws light on the issue.[306]

The Georgia Supreme Court was undoubtedly correct that photographs offered mechanical advantages over human testimony, but its uncritical analogy of the camera to a "dumb witness"[307] without capacity to lie was less persuasive, even in the 1800s.

Indeed, the historic record indicates that observers were aware from the start of the potential for manipulation and deception when photographs were used as evidence.[308] A short and unapologetic critique of photographic evidence appeared in a number of publications in 1886 under the title *The Photograph as a False Witness*.[309] In that article, the anonymous author warns that unguarded acceptance of photographs as legal proof creates a danger of deception and perjury: "[T]he photograph may be made to speak for this or for that, according as the finger of mammon does point."[310] Careful selection of lighting, perspective, and equipment could be used to editorialize the content of a photograph in ways that an unsophisticated audience—or the Georgia Supreme Court—might not suspect.[311]

Postexposure manipulation of photographs was also a concern well before "Photoshop" became a verb. In an 1861 article, Oliver Wendell Holmes, father of the later Supreme Court Justice, quipped, "A simple photographic picture may be tampered with. A lady's portrait has been known to come out of the finishing-artist's room ten years younger than when it left the camera."[312] It seems this type of manipulation was widespread. In one sensational example from the 1860s, photographer

[306] Franklin v. State, 69 Ga. 36, 42–43 (1882) (emphasis added).

[307] *Id.*

[308] Carter, *supra* note 304, at 35–36 ("Staged and manipulated photographs—including photographs that had their negatives retouched, combined, or otherwise tampered with—were widely created and circulated from the very beginning of photographic history, and contemporaries readily understood the artifice employed in the creation of the images.").

[309] *The Photograph as a False Witness*, 30 Photographic News 465 (1886) [hereinafter *The Photograph*]; Photographic News, *The Photograph as a False Witness*, 10 Va. L.J. 644 (1886); Photographic News, *The Photograph as a False Witness*, 34 Alb. L.J. 457 (1886).

[310] *The Photograph*, *supra* note 309, at 465.

[311] *See, e.g.*, *id.* (providing an anecdote in which a lawyer is warned that photography could become a source of deceptive evidence in ancient-lights cases). *See generally* Charles Scott, Photographic Evidence (1st ed. 1942) (illustrating how differences in composition could influence the resulting photographs).

[312] Oliver Wendell Holmes, *Sun-Painting and Sun-Sculpture; With a Stereoscopic Trip Across the Atlantic*, Atl. Monthly, July 1861, at 13, 15 (on file with Atlantic, Print Edition Archive, https://cdn.theatlantic.com/media/archives/1861/07/8-45/131953800.pdf [https://perma.cc/5F9X-XYUW]).

William H. Mumler became the target of popular and legal controversy for his production of spirit photographs—portrait photos which, when developed, appeared to show the spirits of his subjects' deceased relatives floating as ghostly apparitions above them.[313] Whatever technique Mumler used to doctor these photographs was clever enough to evade detection by experienced photographers who visited the studio to observe his process.[314]

The law of evidence eventually settled on handling photographs by analogy to paintings and other constructed representations of witness testimony.[315] The approach and its reasoning are well captured in an early and influential comment on the subject by the New York Court of Appeals:

> A portrait or a miniature taken by a skilled artist, and proven to be an accurate likeness, would be received on a question of the identity or the appearance of a person not producible in court. Photographic pictures do not differ in kind of proof from the pictures of a painter. . . . It is the skill of the operator that takes care of [details like lighting, position, and equipment], as it is the skill of the artist that makes correct drawing of features, and nice mingling of tints, for the portrait. . . . So the signs of the portrait and the photograph, if authenticated by other testimony, may give truthful representations. When shown by such testimony to be correct resemblances of a person, we see not why they may not be shown to the triers of the facts, not as conclusive, but as aids in determining the matter in issue, still being open, like other proofs of identity or similar matter, to rebuttal or doubt.[316]

Put another way, the photograph, like the painting, could be authenticated by a testifying witness as an illustration of that witness's testimony.[317] Somewhere in the background, the photograph still retained its mechanical advantages. But, in the legal theory of the trial, these advantages were set aside as the photo's purpose was merely to help lend color and detail to a witness's spoken words.[318] It was the testimony, not the photograph, that was the evidence before the court.[319] Any risk of deception was thus no different from the traditional risk of

[313] *See* Mnookin, *supra* note 300, at 27–43.

[314] *Id.* at 31.

[315] *Cf. id.* at 53–59 (considering ways in which the treatment of photographs diffused some of the discomfort that judges and jurors might otherwise have felt about factfinding in a context bounded by photographs).

[316] Cowley v. People, 83 N.Y. 464, 477–78 (1881).

[317] *See id.* at 478.

[318] Mnookin, *supra* note 300, at 44.

[319] *See, e.g.*, *id.* at 44–45 (citing late-1800s authority for this understanding of photographs).

false testimony, addressed by existing rules and procedures that policed the accuracy of what witnesses said on the stand.[320]

This limited and rather artificial understanding of photographic evidence survives today as what is sometimes called the "pictorial testimony" use of photographic evidence.[321] In this approach, a photo, video, or similar recording is introduced at trial for the purpose of illustrating a witness's testimony,[322] usually after being authenticated by that witness as a fair and accurate representation of their testimony.[323] There is, in principle, no difference between a candid and a staged photograph in this approach; both are merely illustrations of what the witness is trying to explain. Indeed, photographs introduced only to illustrate a point are commonly said to be "not evidence" at all.[324] Such records have, in theory, no evidentiary weight and would typically not be made available to the jury during deliberations.[325]

At the opposite extreme, another modern use of photographic evidence often goes under the label of the "silent witness" theory.[326] In this approach, a photograph, video, or similar recording is authenticated by a witness with knowledge of its source to be the output of a system that produces reliable results.[327] It may then be introduced as substantive evidence of its content.[328] As one common example, the maintainer of a bank's closed-circuit surveillance-camera system could take the stand to explain how the system works and why its recording of a robbery could be trusted as an accurate depiction of what took place.[329] So authenticated, the recording's probative value would arise

[320] *See id.*

[321] *See, e.g.*, Mooney v. State, 321 A.3d 91, 93–94 (Md. 2024).

[322] *See* 22 Wright & Miller's Federal Practice & Procedure § 5172.4 (2d ed. 2025).

[323] *See, e.g.*, People v. Bowley, 382 P.2d 591, 594 (Cal. 1963) ("It is well settled that the testimony of a person who was present at the time a film was made that it accurately depicts what it purports to show is a legally sufficient foundation for its admission into evidence.").

[324] *See, e.g.*, Fed. R. Evid. 107(b) ("An illustrative aid is not evidence . . . ."); *see also* John Henry Wigmore, Wigmore's Code of Evidence §§ 726–727 (3d ed. 1942) (describing such photographs as "supplement[s]" to and "a part of" testimony).

[325] *See, e.g.*, Fed. R. Evid. 107(b) ("An illustrative aid is not evidence and must not be provided to the jury during deliberations unless: (1) all parties consent; or (2) the court, for good cause, orders otherwise."); *see also* Wright & Miller's Federal Practice & Procedure, *supra* note 322, § 5174 ("[M]ost courts seem to follow the suggestion by the commentators that illustrative objects should not be sent to the jury room during deliberations.").

[326] People v. Bowley, 382 P.2d 591, 594–95 (Cal. 1963) ("[P]hotographs are useful for different purposes. When admitted merely to aid a witness in explaining his testimony they are, as Wigmore states, nothing more than the illustrated testimony of that witness. But they may also be used as probative evidence of what they depict. Used in this manner they take on the status of independent 'silent' witnesses.").

[327] *See* Fed. R. Evid. 901(b)(9).

[328] *See id.*

[329] *See, e.g.*, United States v. Clayton, 643 F.2d 1071, 1073 (5th Cir. 1981) ("[P]hotographs made from bank camera films were sufficiently authenticated by Government witnesses who were

directly from its unthinking, mechanical transcription of the world, not merely from its derivative value in illustrating the firsthand testimony of a human witness.[330] Subject to other relevant rules of evidence,[331] the silent-witness recording could be introduced as substantive evidence itself, essentially as the testimony of the type of "dumb witness" that the Georgia Supreme Court imagined.[332]

The space between photographs as "pictorial testimony" and photographs as "silent witnesses" remains uncomfortably wide. Photographs introduced as pictorial testimony obviously convey gratuitous details beyond the words being uttered by the authenticating witness. Testimony that such a photo is a fair and accurate representation of the scene typically does nothing to establish the value of its fine details, and it is fantasy to believe that jurors interpret such photographs as mere illustrations to be doubted in every respect.[333] At the other extreme, photographs introduced under silent-witness theories may fail to disclose their exposure to human manipulation. Even setting aside more complicated issues, like selection bias when interested parties identify and produce photographic evidence, the susceptibility of images and videos to post-recording manipulation often goes unexplored—a point that has spawned decades of frustrated legal commentary.[334]

not present at the robbery when the testimony adduced stated the manner in which the films were used in the camera, how the camera was activated, that the film was removed immediately after the robbery, and the chain of possession of the film and the development of the prints.").

[330] *E.g.*, United States v. Taylor, 530 F.2d 639, 641–42 (5th Cir. 1976) ("In the case before us it was, of course, impossible for any of the tellers to testify that the film accurately depicted the events as witnessed by them, since the camera was activated only after the bank personnel were locked in the vault. The only testimony offered as foundation for the introduction of the photographs was by government witnesses who were not present during the actual robbery. These witnesses, however, testified as to the manner in which the film was installed in the camera, how the camera was activated, the fact that the film was removed immediately after the robbery, the chain of its possession, and the fact that it was properly developed and contact prints made from it. Under the circumstances of this case, we find that such testimony furnished sufficient authentication for the admission of the contact prints into evidence.").

[331] *See, e.g.*, FED. R. EVID. 1001–1004 (requiring the production of originals or mechanical duplicates of a recording in most such circumstances).

[332] Franklin v. State, 69 Ga. 36, 43 (1882).

[333] *Cf.* Mnookin, *supra* note 300, at 26 ("If the photograph was properly understood as equivalent to any other form of human testimony, then the widespread belief in inherent photographic certainty might make the legal use of this new technology highly misleading.").

[334] *See, e.g.*, WRIGHT & MILLER'S FEDERAL PRACTICE & PROCEDURE, *supra* note 322, § 5172.4 ("[I]t is rare to find a federal court excluding photographic evidence. So far as we can detect, the availability of computer programs that can fake photographs has not made courts any more cautious about admitting photos." (footnotes omitted)); Jill Witkowski, Note, *Can Juries Really Believe What They See? New Foundational Requirements for the Authentication of Digital Images*, 10 WASH. U. J.L. & POL'Y 267, 271–72 (2002) ("Digital images are highly susceptible to manipulation. Manipulation, as distinct from enhancement, consists of changing the elements of a photograph or image by changing the colors, moving items from place to place on the image, or otherwise altering the original image. . . . The electronic nature of the image file makes undetectable manipulation of

As background context for the future treatment of deepfakes, the history of photographic evidence is again a curious mix of causes for concern and comfort. Deepfakes are photographic manipulation carried to its logical extreme. But opportunities for manipulation, deception, and simple overweighing of photographic evidence have existed since the dawn of this technology. Whether and how deepfakes are really all that different from photographs is the subject we next consider.

IV. Deepfakes and Proposed Reforms

Deepfakes are novel. They are shocking. And, as we have already discussed, they are generating a buzz of worried analysis and calls for reform in academic and legislative-policy circles.[335] But deepfakes are also just the newest version of the common lie. We humans have been guarding ourselves against lies and other acts of trickery for a very, very long time.[336] How do deepfakes stand when viewed through the lenses of epistemology and the law of evidence? Are new laws and social interventions as urgent and necessary as they appear to be?

For the most part, we think not. In the following pages, we evaluate common justifications for alarm and corresponding proposals for policy reform. We argue that the case for panic is overstated, and the justifications for reform are insubstantial. We do not deny that deepfakes present new and worrying opportunities for deception in the courtroom. And deception should never be treated lightly—least so in the courtroom. But to accord deepfakes appropriate gravity is not necessarily to treat them differently than other forms of lies and deception. The novel expression of an ancient problem does not necessarily require novel solutions.

Our analysis in this Part draws on justificatory frameworks from both evidentialism and reliabilism. We assume that in the courtroom, factfinders use evidentialist methods to form the beliefs that determine case outcomes. In so doing, we are only taking courts at their word when, for example, they instruct jurors that their "first duty is to decide the facts from the evidence in the case."[337] When we evaluate existing or proposed rules of evidence, we employ a reliabilist point of view. In other words, we assess rules of evidence by how reliably they enable evidentialist jurors to exercise human judgment in arriving at the truth.

Our conclusion is that knee-jerk proposals in the literature tend to focus on deepfakes as isolated pieces of evidence. They mistakenly

a digital image easy, in part because no traditional 'original image' is made. Unlike traditional cameras, which produce one negative, digital cameras create an electronic file from which the image can be generated.").

[335] *See supra* Section I.C.

[336] *See supra* Parts II–III.

[337] *E.g.*, Fed. Civ. Jury Instructions of the Seventh Cir., § 1.01 (2017).

assume that deepfakes will always bear some mark of their false provenance. Or they forget that, like any piece of evidence, digital media need not bear their own mark of authenticity to be deemed trustworthy. But all evidence is situated within a web of codependencies, and the law has long relied on human judgment about context to help disentangle fact from fiction.

Before turning to the arguments, one clarification may be helpful. Our focus, here, is on the use of deepfakes to deceive—that is, generated media being presented to the judge and jury as if they were simple, mechanical recordings of reality. This limited scope of analysis is important because different issues are raised by something like the clearly disclosed use of computer-generated content to illustrate a witness's testimony—what we would call "deep fabrications." "Deepfabs" are interesting in their own right, but they are not our focus in this Article. Different issues are likewise raised by the autonomous editing decisions of smart devices. For example, smartphones use filters, exposure settings, and postprocessing to convert raw recordings of nighttime scenes into clear and attractive photographs.[338] This type of transparent background editing by "silent *smart* witnesses" presents interesting evidentiary and epistemological challenges. But these, too, are not our focus in this Article.

A. Conduct-Oriented Prohibitions and Penalties

Turning to deepfakes as deception and corresponding proposals for reform, we can start with what might seem like the most targeted responses to the challenge: proposals that would prohibit deepfakes from being produced and distributed in the first place. Examples include calls for compelled origin-disclosure statements on generated media[339] and calls to ban and penalize specific abuses of deepfakes, like the production of a video portraying a targeted person engaging in a sex act.[340] At the extreme, this strategy could be implemented as a flat ban on the production and distribution of *any* deepfake content.[341] Reframed

[338] *See, e.g.*, *Take Great Photos and Videos*, Apple, https://support.apple.com/guide/iphone/take-great-photos-and-videos-iph9bbc8619e/ios [https://perma.cc/Z76J-8JVA] (last visited Nov. 5, 2025) ("Night mode automatically improves photos taken in low light on supported iPhone models.").

[339] *See* Delfino, *supra* note 12, at 303 (describing an act that would have "mandated that most classes of deepfakes" conspicuously disclose their fabrication, with penalties available to enforce this requirement).

[340] *See* Brown, *supra* note 18, at 45–47 (describing state law proposals for banning the use of deepfakes in attempting to influence elections and in generating sexually explicit content without consent).

[341] *Cf.* Chesney & Citron, *supra* note 20, at 1788–89 ("[A] flat ban is not desirable because digital manipulation is not inherently problematic.").

in epistemological terms, these proposals evince confidence in the reliability of existing court rules but worry about the future justificatory power of digital media as deepfakes dampen the truth signal digital media provides to jurors. Banning deepfakes boosts the signal of digital media that remain, making it easier for jurors to form justified beliefs based on it—or so the reasoning apparently goes.

Proponents of bans on deepfakes may appropriately aspire to address more than our specific focus on the deceptive use of deepfakes as trial evidence.[342] But if their proposals are to address the courtroom challenge, then they should at least provide some identifiable advantages over existing rules of evidence and related restrictions. For this to happen, two conditions must be satisfied. First, the existing legal safeguards must be inadequate to deter the introduction of deepfakes into evidence. Second, the proposed bans must offer credible improvements in deterrence over existing law.

The first of these conditions is almost surely satisfied. True, there are many deterrents to presenting false evidence in the courtroom. A lawyer cannot ethically mislead a court or facilitate the presentation of evidence that the lawyer knows or reasonably believes to be untrue.[343] Every witness who testifies must first "give an oath or affirmation to testify truthfully."[344] Because no purported recording can be introduced as evidence without being authenticated as accurate by the testimony of a witness with appropriate knowledge of its accuracy, known deepfakes cannot be introduced without someone lying to the tribunal and thus subjecting themselves to the penalty of perjury.[345] But the actual enforcement of penalties for perjury is infrequent at best,[346] and almost the entire history of the law of evidence betrays the commonsense understanding that promising to tell the truth is little obstacle to lying.[347]

The second condition is where the proposed reforms fall flat. If the oath and all related penalties for lying in court are not already adequate to prevent deepfakes from being presented as legitimate evidence, what

[342] *See id.* at 1771–86 (describing social, political, and other problems that could be caused by the proliferation of deepfakes).

[343] *See, e.g.*, Model Rules of Pro. Conduct r. 3.3 (A.B.A. 2023) ("A lawyer shall not knowingly . . . offer evidence that the lawyer knows to be false. If a lawyer, the lawyer's client, or a witness called by the lawyer, has offered material 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 the testimony of a defendant in a criminal matter, that the lawyer reasonably believes is false.").

[344] Fed. R. Evid. 603.

[345] *See* Charles Doyle, Cong. Rsch. Serv., 98-808, False Statements and Perjury: An Overview of Federal Criminal Law 13–14 (2024).

[346] *See* Comment, *Perjury: The Forgotten Offense*, 65 J. Crim. L. & Criminology 361, 361 (1974). *But cf.* Chris William Sanchirico, *Evidence Tampering*, 53 Duke L.J. 1215, 1224 (2004) (presenting restrictions on lies and evidence tampering in a more optimistic light).

[347] *See supra* Section III.A.

contribution does one more rule against deception stand to make? Unless proposed legislation offers greater or more certain penalties for deepfake deception than for other examples of lying under oath, the promises of additional penalties are hard to spot.

For deepfake bans to have any additional deterrent effect, deepfakes must also be at least reasonably detectable. How else would production and promulgation be punishable except if the result was identifiably fake upon inspection? As we have already discussed, deepfake detection is an active area of research,[348] but the arms race between deepfake detectors and generators looks unpromising for detectors.[349] For early deepfakes and crude manipulations, conduct-oriented bans may perhaps do some work. But in a world of undetectable deepest fakes, these interventions are toothless.

B. Prophylactic Exclusionary Rules

If conduct-oriented prohibitions cannot stem the predicted tide of deepest fakes, then the law of evidence is the next logical place to look for solutions. Here, there is an immediate positive note. Increasingly searching scrutiny of evidence presented as standard mechanical recordings seems likely to be an organic byproduct of the adversarial system's growing awareness of deepfake technology.[350]

The reason for this is simply that the authentication standard is a fact question embedded in the changing social context. In order to introduce photos, videos, and other recordings into evidence, the proponent must be able to defend the authenticity of the evidence as being what the proponent claims it to be.[351] The proponent must also persuade the factfinder to give that evidence whatever weight it deserves.[352] It takes little imagination to see why opposing counsel, in a world in which deepfakes are plentiful, would be more apt to challenge the authenticity of apparently recorded evidence than they are today.[353]

In litigating these challenges, proponents of deepfake-able evidence are also likely to be chasing increasingly demanding targets. To see why, consider the lowly authentication standard, usually articulated as requiring "evidence sufficient to support a finding [by a

[348] *See supra* Section I.C.

[349] *See supra* Section I.D.

[350] *See generally* Allen et al., *supra* note 283 (discussing the role of the adversarial process in addressing many evidentiary challenges).

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

[352] *See* Delfino, *supra* note 12, at 321.

[353] *See* Pfefferkorn, *supra* note 28, at 268 (predicting more frequent litigation of authenticity to mean that "successfully getting a video admitted into evidence may require additional motion practice, witness testimony, and forensic tools").

preponderance of the evidence] that the item is what the proponent claims it is."[354] Now consider this standard in relation to an audio file presented as a recording of the defendant's verbal confession. In a world without deepfake-voice generation, this audio file could be convincingly authenticated by simple means. The jury could compare the recorded voice to that of the defendant in deciding that the recording probably was the defendant's spoken words.[355] But in a world of deepest fakes, that simple demonstration may fail to persuade. Even if the evidence is admitted, the jury may assign it little weight out of fear that it could have been artificially generated by the prosecution. Authenticity and persuasion are both context-dependent requirements, and as the ease of producing deepfakes increases, it is only natural to suppose that factfinders will grow increasingly skeptical of the purportedly recorded evidence put before them.

Although some commentators fear the realization of this prediction—a reaction we take up in the next Section—others demand more than what organic change promises to produce. These commentators propose to change the law of evidence to reduce the opportunities for deepfake deception.[356] One version of this proposal would withdraw the silent-witness theory of authentication altogether.[357] Broader yet, evidence law could be changed to prophylactically exclude every type of deepfake-able evidence, perhaps on the reasoning that it is impossible to demonstrate that such evidence is conditionally relevant.[358]

In epistemological terms, these proposals sound more in the vein of digital media skepticism. If we cannot prevent deepest fakes from proliferating, the thinking goes, then digital media will eventually carry a very low truth signal. As a consequence, jurors will not be able to form justified beliefs on the basis of digital media evidence. To maintain courts' reliability as a truth-finding process, digital media must be excluded from trial or its use severely limited.

Here, again, we see these proposals as poorly calibrated to the challenge they purport to address. The problem is not that they are

[354] Fed. R. Evid. 901(a).

[355] *See* Fed. R. Evid. 901(b)(5).

[356] *E.g.*, Delfino, *supra* note 12, at 297 ("The current Rules [of Evidence] will need to be adapted to solve the problem of how to show when a video is fake and when it is not."); *id.* at 332 ("Standing alone, none of the Federal Rules of Evidence or their companion common-law theories are sufficient to address the significant challenges that deepfakes present . . . .").

[357] *See id.* at 341 ("[T]he silent witness theory will not be helpful when handling deepfakes, because the technology is too sophisticated to warrant the trust required to authenticate evidence under this theory without an authenticating witness."); *see also* Danielle C. Breen, *Silent No More: How Deepfakes Will Force Courts to Reconsider Video Admission Standards*, 21 J. High Tech. L. 122, 160 (2021) ("Absent significant deepfake legislation, courts should adopt the pictorial evidence theory to combat heightened public skepticism of photographic and video evidence.").

[358] *See* Fed. R. Evid. 104(b).

necessarily misguided, but that they overstate the severity of the deepfake threat—and even the deepest-fake threat. They do this by failing to account for context in how the evidence will be assessed as authentic or fake.

To illustrate, imagine the trial of a civil action arising from a car collision at an intersection. The plaintiff wishes to introduce a video recording that purportedly shows the light was green as the plaintiff's car entered the intersection. This video was shot on the smartphone of a disinterested third-party witness. This third party was trying to record a video of her dog doing a trick but accidentally caught footage of the collision in the background. The witness takes the stand and testifies that she did not observe the collision when it happened because her attention was on the dog, but she is sure that the video was made using default settings on her phone. She observes the video and testifies that it looks today exactly as it did when she filmed it. She also produces her phone for inspection; the recording, still present in her photo application, is identical in every way to the video file that the plaintiff seeks to introduce as evidence.

Our question for digital media skeptics is this: Is the mere technical feasibility of deepfake-video generation sufficient ground for excluding the video evidence in this hypothetical? Just as the conceptual possibility of Descartes's demon is insufficient to justify external-world skepticism, we think the answer should be an emphatic "No." True, the scene could have been generated from nothing more than a text prompt. But why would a disinterested third party generate a false video, save it to their smartphone, and then lie about the doctored origin of that video under oath, all in relation to an unconnected legal dispute? Broadening the reasoning beyond this example, why should the technical feasibility of deepfake-content generation pose a problem for introducing *any* evidence that is credibly authenticated by disinterested witnesses?

A useful way of reframing proposals to exclude deepfake-able evidence is to note their similarity to an extreme version of the old competency rules for witness testimony. As we have previously discussed, trial practice once struck witnesses from the stand in contexts in which they were not expected to be truthful despite their promise not to lie.[359] Prophylactic exclusionary rules for deepfakes would operate the same way and for essentially the same reason: Because digital media can now deceive, courts must exclude it all. This framing also helps to illustrate the problem with the recent proposals. Not even the harshest competency rules excluded *all* witnesses from giving testimony. Witness testimony was by default admissible and excludable only in circumstances in which lying was a live concern.

[359] *See supra* notes 262–71 and accompanying text.

We would not go so far as to endorse anything like the old competency rules for addressing the problem of deepfake evidence. The deficiencies of the competency system were not limited to its dubious utility in helping juries resolve hard cases.[360] But if prophylactic exclusion is ever deemed a worthy intervention to pursue, it should not exclude all media. Like the competency rules of old, it should be limited to evidence produced by witnesses with a plausible motivation to lie under oath.

C. *The "Deepfake Defense" and Juror Skepticism*

Finally, recent discussions about deepfakes and their role in trials have raised alarm about the possibility of a blanket deepfake defense that could be lobbed against even genuine and accurate media evidence.[361] A related line of reasoning worries that digital media skepticism will overtake jurors in a world in which deepfakes are everywhere. For the most part, these concerns are presented as systemic worries about how deepfakes will change the way that factfinders process evidence. Of a more functional mindset, one recent proposal attempts to tackle the deepfake defense and juror skepticism by amending the rules of evidence to concentrate authentication decisions in the hands of judges.[362] Under the proposed rule, the judge who finds a proffered item of evidence authentic would instruct the jury not to doubt its authenticity.[363] Epistemologically, such proposals worry that digital media skepticism is inevitable outside of court. Such proposals, however, believe that more robust gatekeeping in the courtroom could preserve the truth signal digital media sends at trial.[364]

We understand the bases for predicting increased invocation of the deepfake defense and rising juror skepticism, but we struggle to understand why either of these is a problem. In considering the deepfake

[360] *See* Fisher, *supra* note 257, at 659–97 (discussing these deficiencies); *see also* Langbein, *supra* note 258, at 1185 (describing even "disqualification for interest" as "a grievous shortcoming in common law civil procedure").

[361] *E.g.*, Delfino, *supra* note 12, at 310 ("This 'deepfake defense' will debut in court in the foreseeable future, if it has not already."); Pfefferkorn, *supra* note 28, at 255 ("The opponent of an authentic video may allege that it is a deepfake in order to try to exclude it from evidence or at least sow doubt in the jury's minds."); Chesney & Citron, *supra* note 20, at 1785 ("As the public becomes more aware of the idea that video and audio can be convincingly faked, some will try to escape accountability for their actions by denouncing authentic video and audio as deep fakes. Put simply: a skeptical public will be primed to doubt the authenticity of real audio and video evidence. . . . Hence what we call the liar's dividend: this dividend flows, perversely, in proportion to success in educating the public about the dangers of deep fakes.").

[362] *See* Delfino, *supra* note 12, at 341–42.

[363] *See id.* at 342 ("The court would . . . admonish the jury to weigh that evidence, but not question its authenticity.").

[364] *See id.* at 348.

defense, it cannot be forgotten that *every* item of evidence is vulnerable to attack for its lack of authenticity, accuracy, and reliability. True, one party can now claim that the other party's evidence is a deepfake.[365] But what is the significance of this claim if credible evidence is not available to back it up? Return, again, to the third-party auto-collision footage. Would a claim that the video is fake, based on nothing but theoretical possibility of deepfake generation, be worthy of deep analysis? Courts that have responded to this question to date offer an answer that largely mirrors our own: The conceptual possibility of fabrication alone is insufficient to support an authenticity challenge.[366]

But what about runaway juror skepticism? Suppose the overwhelming spread of deepfakes in everyday life pushes jurors to the point that they doubt the accuracy of all media evidence. We concede that this is a dystopian vision of the world. But we, again, fail to see why it is a problem. Trials nearly always present jurors with conflicting evidence of variable quality, and some of questionable sincerity. Jurors are expected to bring their common sense, their life experience, and their own healthy skepticism to the task of evaluating the evidence they are presented with at trial. If the community at large comes to distrust media evidence, then that distrust can and should make its way into jurors' trial deliberations. Far from a problem to be corrected, this is the system working as intended.

Conclusion

What lessons does our tour through epistemology and evidence law hold for how society will respond to deepfakes? Alarmists warn of an epistemic apocalypse that will erode not only our ability to know true things, but perhaps our very concept of truth.[367] Law scholars in

[365] *See* Audrey Mitchell, *Deepfaked Evidence: What Case Law Tells Us About How the Rules of Authenticity Needs to Change*, Berkeley Tech. L.J.: Blog (June 23, 2025), https://btlj.org/2025/06/deepfaked-evidence-what-case-law-tells-us-about-how-the-rules-of-authenticity-needs-to-change/ [https://perma.cc/5RS6-D8G7] (providing examples of cases in which a party has alleged that evidence is a deepfake).

[366] *See, e.g.*, People v. Foreman, 2020 IL App (2d) 180178-U, ¶ 145 ("Defendant . . . points out that improperly-authenticated recordings are inherently suspect in this age of deep-fake videos and easily-manipulated audio records. We reject defendant's arguments. . . . We also reject defendant's argument that recent technological advancements render all recordings suspect, because they can be easily manipulated. In the absence of any evidence of tampering or other such manipulation in this case, there are no foundational issues with the recordings."); Pittman v. Commonwealth, No. 0681-22-1, 2023 WL 3061782, at *6 (Va. Ct. App. Apr. 25, 2023) ("[T]here is no evidence of or contention that would call into question the veracity of the video or the possibility of a 'deep fake.' And we reiterate that where there is 'mere speculation that contamination or tampering could have occurred, it is not an abuse of discretion to admit the evidence and let what doubt there may be go to the weight to be given the evidence.'" (quoting Reedy v. Commonwealth, 388 S.E.2d 650, 652 (Va. 1990))).

[367] *See, e.g.*, Fallis, *supra* note 181, at 624–27.

this camp predict that our adjudicatory practices will fare no better. Because "[t]he courtroom is a microcosm of society in general,"[368] people bring their everyday frameworks and assumptions with them when they become judges and jurors. If a post-truth dystopia reigns on the outside, it must reign in the courtroom, too.

In this concluding Part, we reverse the courtroom-society analogy. If courtrooms mirror society, we can also look to adjudication for insights that could unfold in ordinary life. Since the beginning of modern evidentiary practice, courts have grappled with the fact that most evidence is created by humans, and human creations can be tools for deception. The duplicitous twins of testimony, documents, and photographs are perjury, forgeries, and fauxtographs. As discussed above, the predecessors of deepfake alarmists were equally concerned that each new type of evidence would irreversibly corrupt courts' truth-finding function.[369] These alarmists defended varieties of philosophical skepticism about whether the artifacts they considered could ever facilitate truth finding in court.

With time, every one of these initially alarming sources of concern has faded. Today, testimony, documents, and photographs are alive and well as fixtures of evidentiary practice. This is not because alarmists were wrong about the nature of these artifacts. Since the 1860s, photographers have been capable of generating false images.[370] The capacity to lie is even older and requires no technical expertise.[371] The moral that courts repeatedly affirm and that alarmists seem periodically to forget is that human epistemic judgment is nuanced, flexible, and crucially multivariate. We look beyond the four corners of a statement, document, or photograph to evaluate its truth. We can catch out a lie not because we are astute lie detectors—we are not[372]—but because we recognize that the statement is embedded within a web of other evidence, contexts, and commonsense intuitions that bear on its veracity. Does the photograph match the credible testimony of witnesses? Does it correspond with other photographs of the subject?[373] Does the context in which it

[368] Pfefferkorn, *supra* note 28, at 257.

[369] *See supra* Section III.B.

[370] *See supra* notes 313–14 and accompanying text.

[371] *See* Letter from Mark Twain to Margery H. Clinton (Aug. 18, 1908) (commenting that the expression "Let a sleeping dog lie" is "a poor old maxim, & nothing in it: anybody can do it, you don't have to employ a dog"), *reprinted by* TWAIN'S GEOGRAPHY, https://twainsgeography.com/node/11416 [https://perma.cc/5TJM-82CS].

[372] *See supra* note 277 and accompanying text.

[373] A fascinating example of this mode of assessment is found in an early comment by Oliver Wendell Holmes, presenting stereoscopic recording as more reliable than photographs due to the difficulty of simultaneously manipulating the stereograph's dual images:

> [T]ry to mend a stereograph and you will soon find the difference. Your marks and patches float above the picture and never identify themselves with it. . . . The impossibility of

was produced lend support or suspicion to its contents? When we are concentrated on finding the truth, the epistemic weight we attach to every item of evidence constantly evolves in light of other evidence and our own experience as we learn more and more about the subject at hand.

Courts' historical solution to falsifiable evidence has not been less evidence, but more. This, we predict, will be their response to deepfakes too. The existence of deepest fakes does not mean that jurors will be left with coin tosses and guesses when presented with digital media evidence. Rather, jurors will simply require more before believing and will become more astute judges of the fuller array of considerations that bear on truth.

Before envisioning how deepfakes will play out beyond the courtroom, consider a related example from recent memory. In the not-too-distant past, people could generally trust the claims of disinterested third parties. Where is the best burger in town? What is a home remedy to whiten my teeth? Where was Kamala Harris born? There was never a guarantee that the response of a random passerby would be true, but you could generally trust them to provide their best answer, if they answered at all. As discussed above, this trust is grounded in the generally accepted norms of sincerity and competency that govern the ethics of interpersonal testimony.[374] The internet changed this or at least limited its scope. The information superhighway is not the country road. Trolls were born in the faceless corners of the web. Unchecked by any reputational consequences for violating the norms of sincerity and competence, they asserted falsehoods simply to bait clicks or stoke response. Outdated testimonial expectations could misguide unwary netizens to believe what the trolls wrote. Many an uncle has pounded the Thanksgiving table, swearing by some preposterous claim he read in some unremembered online forum.

But the troll did not kill testimony. Rather, society updated its epistemic norms to weaken the prima facie evidentiary value accorded to online statements—the brute fact that some text appears on the internet is little reason to believe what it says. Now everyone knows you cannot believe everything you read on the internet. Before believing it, people must exercise judgment that contextualizes the statement. It is relevant, for example, who made the claim, what their interests are, where the statement appears, how it connects to other facts, whether it conforms to common sense, etc. The trolls are still out there, and their

> the stereograph's perjuring itself is a curious illustration of the law of evidence. "At the mouth of *two witnesses*, or of three, shall he that is worthy of death be put to death; but at the mouth of one he shall not be put to death."

Holmes, *supra* note 312, at 15.

[374] *See supra* notes 196–98 and accompanying text.

knowing falsehoods are still indistinguishable on their face from everyone else's attempted truths. But the trolls' epistemic power diminished as we became more astute consumers of online text.

The same natural cycle of epistemic updating will play out as society emerges from the present alarm over deepfakes. Just five years ago, people could generally trust what they saw in videos. Of course, with big budgets, lots of time, and sufficient expertise, highly motivated actors could use computer-generated imagery to create persuasive fakes. Because most videos did not and could not implicate these concerns, the signal videos sent to people searching for truth was fairly high. Deepfakes changed this, and as their prevalence increases, the average signal that videos send will weaken. The bumpy period is now as the number of deepfakes grows faster than the public's awareness of them. The signal is weaker, but not everyone knows it. This, of course, is rapidly changing. Searches of English-language corpora show that references to deepfakes are doubling biennially,[375] and references in news outlets are currently doubling annually.[376] Growing awareness of deepfakes will "alter our trust in audio and video for good."[377] We anticipate the imminent rise of the dinner table meme, "You can't trust everything you see in a video."

That is where the alarmists end the story, but it is not the story's end. You can see their error most clearly in statements like "[i]f viewers cannot distinguish authentic videos from fabricated ones on their own, they will be disinclined to trust *any* video."[378] Not trusting everything is a far cry from trusting nothing. As jurors have done for testimony, documents, and photographs, and as society has recently done for text online, we will learn to use contextual factors to discern high-signal from low-signal content.[379] The alarmists are right that society will become "a skeptical public . . . primed to doubt the authenticity of real audio and video evidence."[380] But they are wrong that this means people will

[375] Google Books Ngram Viewer, "deepfake" (Oct. 21, 2025), https://books.google.com/ngrams/graph?content=deepfake&year_start=2015&year_end=2022&corpus=en&smoothing=3&case_ insensitive=false [https://perma .cc/MW5X-2TMQ].

[376] Mark Davies, *NOW Corpus (News on the Web)*, English-Corpora, https://www.english-corpora.org/now/ [https://perma.cc/4Y5Z-47M6] (last visited Nov. 29, 2025) (select "chart"; then search in search bar for "deepfake").

[377] Donie O'Sullivan, *When Seeing Is No Longer Believing*, CNN Bus., https://www.cnn.com/interactive/2019/01/business/pentagons-race-against-deepfakes/ [https://perma.cc/F63K-MRR8] (last visited Nov. 5, 2025).

[378] Brown, *supra* note 18, at 11.

[379] *See* Anne E. Boustead & Matthew B. Kugler, *Juror Interpretations of Metadata and Content Information: Implications for the Going Dark Debate*, J. Cybersecurity, Feb. 21, 2023, at 1, 2, 8–9.

[380] Chesney & Citron, *supra* note 20, at 1785; *see* Delfino, *supra* note 21, at 1082 ("As public knowledge of deepfakes continues to grow . . . people [will] become increasingly skeptical about the credibility of audiovisual images . . . .").

become skeptics. People will simply become more astute consumers of digital media.

More concretely, this means the human element will become more important for truth finding, not less. Rather than moving passively from seeing to believing (the old model) or from seeing to disbelieving (the alarmist prediction), consumers will intervene in their own belief-forming processes with increasingly refined judgment. They will pause to look beyond the four corners of their video players to the sort of factors that digital media literacy advocates[381] and scam advisories[382] have long promoted. Is the content too good, bad, or bizarre to be true? Are the stakes high? Am I presently in a calm state of mind? Where am I accessing the video?[383] Is the source trustworthy? Is the video asking me to do anything? Do other sources confirm the content? None of these questions, in isolation or combination, can provide certainty. But as we come to ask them, we will become equally discriminating consumers of video as we already are of testimony and text. In a sense, deepfakes will be an epistemic boon rather than the epistemic harm philosophers fear. They will force us to become better believers.

[381] *See Digital Media Literacy Fundamentals*, MEDIASMARTS, https://mediasmarts.ca/digital-media-literacy/general-information/digital-media-literacy-fundamentals [https://perma.cc/P2FG-7DW3] (last visited Oct. 21, 2025).

[382] *See, e.g.*, *How to Spot a Scam*, OFF. MINN. ATT'Y GEN., https://www.ag.state.mn.us/consumer/publications/howtospotascam.asp [https://perma.cc/5G5B-CCA9] (last visited Oct. 21, 2025).

[383] "You don't have to become a detective. You don't have to become a forensic analyst . . . . Just get off of social media. You will thank me." Bill Chappell, *LA's Wildfires Prompted a Rash of Fake Images. Here's Why*, NPR (Jan. 16, 2025, at 14:00 ET), https://www.npr.org/2025/01/16/nx-s1-5259629/la-wildfires-fake-images [https://perma.cc/B5N6-XK3J].