Winning the Race: America's AI Action Plan*

Part 2: Combating Malicious Deepfakes and Imaginary Evidence – Has Anybody Seen the Ghost of Cotton Mather Lately?

~ Plus ~

Part 3: Removal Procedures And Criminal Penalties For Nonconsensual Intimate Visual Depictions: An Introduction To The TAKE IT DOWN Act, Its Applications, Mandatory Notice And Takedown Provisions And Criminal Sanctions; With Ancillary Musings On Nonconsensual Intimate Audio Works And Allegedly Injurious Large Language Models – Should We Be Talking To A Libertine Chatbot About This?

By

Gary Rinkerman**

The following two-part discussion highlights two important developments in recent efforts to address the uses of Artificial Intelligence in creating fictitious and damaging content. The topics are, respectively, the creation and detection of false evidence in adjudicative and deliberative processes; and (2) the proliferation of injurious content on the Internet, assisted through the use of Artificial Intelligence. The issues are grouped together because they share some common technological concerns – and they are noted together in the recent White House document, *Winning the Race:* AMERICA'S AI ACTION PLAN. The style of treatment, as with other parts of this series, includes some historical context, meditation and (hopefully) constructive wandering, treatment of related issues, and focus on the current and potential future state of pertinent law and processes. In many ways, the style departs from the usual forms of legal discussion. However, this series of discussions on Artificial Intelligence was conceived as an alternative to the proliferation of articles in more traditional formats.

The core conclusions with regard to the topics treated in the following discussions are: (1) we need a readily available, competitively-checked deepfake and digital fraud detection resource for use in conjunction with evidentiary rules in their current or amended form; (2) the TAKE IT DOWN Act's provisions, including takedown procedures for injurious visual content, should be expanded to expressly include audio content; and (3) the model provided by the TAKE IT DOWN Act's expedited, bi-partisan consideration should now be focused on the roles of Artificial Intelligence in creating potentially dangerous Large Language Models services or products, such as chatbots. These points are placed in context in the relevant sections of the discussions, but other less prominent issues – such as the appropriateness of criminal sanctions provided by the TAKE IT DOWN Act and the deadline to institute notice and takedown procedures – are also considered.

I. Introduction

And this is artificial moonlight, an artificial sky . . . Never seen this picture before.[1]

The potential benefits and the threats posed by rapidly developing artificial intelligence ("AI") technologies and their implementation have attracted attention at virtually all levels of government. This attention is predicated, in part, on the general recognition of the developing capabilities of AI and the consequent need to continually assess and avoid risks that can arise from improper reliance on, defects in, or misuse of AI and its outputs. For example, according to the AI system, ChatGPT, "Government regulation of AI helps ensure it is **safe, fair, transparent, and beneficial for society**, rather than harmful or exploitative. Like with

[1]*This two-part article combines the second and third installments in a series of discussions of *Winning the Race*: AMERICA'S AI ACTION PLAN ("AI Action Plan"), issued by the Trump Administration on July 23, 2025. The two main subjects treated in this combined discussion – synthetic media evidentiary considerations and the TAKE IT DOWN Act – were noted in the same section of the AI Action Plan (although evidentiary considerations predominate) and also appear side-by-side in this two-part discussion. The first article in the series is *Winning the Race: America's AI Action Plan, Part 1: Scrubbing Inappropriate AI Bias – Where Have All The Lysergic Vikings Gone?, See*
https://crc.gmu.edu/winning-the-race-americas-ai-action-plan-article-by-gary-rinkerman/ or
https://care.gmu.edu/winning-the-race-americas-ai-action-plan-article-by-gary-rinkerman/.

**Gary Rinkerman is a Founding Partner at the law firm of Pierson Ferdinand, LLP, an Honorary Professor of Intellectual Property Law at Queen Mary University School of Law in London, a member of George Mason University's Center For Assurance Research and Engineering, and a Senior Fellow at George Mason University's Center for Excellence in Government Cybersecurity Risk Management and Resilience. The views and information provided in this article are solely the work of the author and do not comprise legal advice. They are not for attribution to any entity represented by the author or with which he is affiliated or is a member – including, *e.g.,* the firm in which he is a member or any of its clients. All Internet citations and links in this article were visited and validated on September 4, 2025.

Lyric excerpt from *A Gentleman's Honor*, *The Photographer*, Chamber Opera, Music by Philip Glass; Lyrics based on, excerpted, and arranged by David Byrne from biographical information regarding motion photography pioneer Eadweard Muybridge and from testimony and evidence presented at his 1875 trial for murdering his wife's lover. *See* Philip Glass Interviewed by Russ Jennings, Publication Date: 1983-04-18 (audio recording – time marker approx. 5:00 – 7:22), https://archive.org/details/PhillipGlassWithJennings/phillip_glass_w_russ_je_No41A.wav; *see also,* Leslie, *The Man Who Stopped Time*, Stanford Magazine, May/June 2001, https://stanfordmag.org/contents/the-man-who-stopped-time. The reference to the picture in the quoted excerpt from the lyrics is likely a reference to an annotated authentic photograph that Muybridge inadvertently discovered; the photograph led him to recognize that his wife was having an affair and that his child was likely not his biological child – and that evidence, as the jury determined, was among the critical factors that drove Muybridge to commit an act of "justifiable homicide." Parenthetically, there is a great statue of Muybridge at the Letterman Digital Arts Center in the Presidio, San Francisco – it's worth a visit.

electricity, cars, or the internet, thoughtful rules are key to managing risks while supporting innovation." (emphasis by ChatGPT).[2]

On July 23, 2025, the Trump Administration released *Winning the Race:* AMERICA'S AI ACTION PLAN (the "AI Action Plan").[3] As stated in the release documents, the Plan identifies over 90 Federal policy actions that the Trump Administration will take in the coming weeks and months.[4] These actions will be undertaken in three "pillars" of priority and concern: (1) Accelerating Innovation, (2) Building American AI Infrastructure, and (3) Leading in International Diplomacy and Security.[5]

One of the actions promoted in the AI Action Plan appears under the Section titled: *Combat Synthetic Media in the Legal System*[6] - "synthetic media" being video, images, text, voice, or any combination of them, that has been fully or partially generated by AI.[7] Whether every

---

[2] On August 1, 2025, the author asked ChatGPT "why should the government regulate artificial intelligence?" The response "summary" is provided in the above text. However, the full response is as follows: "Governments should regulate artificial intelligence (AI) for several important reasons. Here are the main ones: **1. Public Safety and Security (a) Prevent misuse**: (a) AI can be weaponized or used for cyberattacks, surveillance, or deepfakes; (b) **Avoid accidents**: In critical applications like self-driving cars or medical diagnostics, poorly regulated AI can lead to life-threatening errors; **2. Ethical Standards and Human Rights: (a) Bias and discrimination**: AI systems can reflect or amplify societal biases (*e.g.,* in hiring, policing, lending); (b) **Privacy protection**: Without regulation, AI may collect and exploit personal data without consent; **3. Economic Fairness: (a) Monopoly risks**: Big tech companies can dominate AI development, stifling competition; (b) **Job displacement**: Automation may replace many jobs—regulation can help manage transitions and support affected workers; **4. Transparency and Accountability: (a) Explainability**: People affected by AI decisions (*e.g.,* being denied a loan) should know why; (b) **Accountability**: Laws are needed to determine who is responsible when AI causes harm or makes bad decisions; **5. National and Global Stability: (a) AI arms race**: Regulation helps prevent a global race to build powerful AI without safety checks; (b) **Standard setting**: International cooperation on AI standards can help avoid conflict and promote beneficial uses; **6. Trust and Public Confidence: (a)** Clear rules make AI development more trustworthy; (b) Consumers and citizens are more likely to accept AI when they believe it's safe, fair, and well-governed. **Summary:** Government regulation of AI helps ensure it is **safe, fair, transparent, and beneficial for society**, rather than harmful or exploitative. Like with electricity, cars, or the internet, thoughtful rules are key to managing risks while supporting innovation. Would you like a short version for a speech or essay? Ask ChatGPT." (emphasis in original; punctuation and formatting modified by this article's author.) Yes, the reliance here on AI was intended to be ironic and amusing, although it should be clear that comedy is not the author's strong suit.

[3] *See* https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf. A precursor statement to the AI Action Plan can be found in President Trump's Executive Order 14179 of January 23, 2025 (Removing Barriers to American Leadership in Artificial Intelligence) which includes provisions revoking the Biden Administration's Executive Order 14110 of October 30, 2023 (Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence). The revocation is based, in part, on the perception that Executive Order 14110 encouraged bias and inappropriate social engineering aspirations in the development and employment of government AI. Trump's Executive Order 141179 states that US leadership in AI innovation requires that "we must develop AI systems that are free from ideological bias or engineered social agendas." *See* Fed. Reg. Vol. 90, No. 20 Friday, January 31, 2025, pp. 8741-8742. *See* https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence.

[4] *See* https://www.whitehouse.gov/articles/2025/07/white-house-unveils-americas-ai-action-plan/.

[5] *Id.*

[6] Pillar 1: Accelerate AI Innovation, *Combat Synthetic Media in the Legal System*, AI Action Plan, pp. 12-13, https://whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

[7] Although there can be overlap, "Synthetic Media" or "Deepfakes" differ, in the general understanding of the terms, from "Machine-Generated Proof," which is used to describe proof provided by machines – such as "Google Earth

potential use of machine generated evidence in the legal system needs to be combatted is open to debate,[8] but there is no debate regarding the need to avoid and counter the use of falsified AI-assisted evidence that corrupts the legal system and its outcomes. Therefore, the AI Action Plan's focus, and the accompanying Recommended Policy Actions, are premised on the need to ensure that federal evidentiary standards and resources address the increasing sophistication of AI-generated deepfakes [9] and their potential uses to cause damage, pervert justice in individual cases, and to corrupt the general operations of our legal system and government.[10] As noted in *Kohn v. Ellison*, the challenges grow as AI deepfake technology becomes more readily available to a broad spectrum of users.[11] As the *Kohn* Court stated:

> The technology used to manipulate images, sounds, and videos existed before deepfakes entered the picture, but such technology was expensive, time-consuming, technical, and difficult to use for the ordinary person. However, in the last 15 years, deepfake technology has become increasingly accessible and available, and ordinary people now can generate convincing deepfake content at little cost in a matter of minutes. And with the concomitant rise of social media, deepfake can be disseminated to millions of online users with the mere click of a button. (citations omitted)[12]

---

location estimates, conclusions of machine-learning algorithms as to attribution of authorship, results provided by forensic software for face and voice recognition, Find My iPhone features used to track phone theft – that is not generated with deception as the motive. Nonetheless, issues of reliability regarding Machine-Generated Proof can arise when, for example, trade secret claims are used to shelter underlying algorithms from scrutiny. *See* Roth, *Proposal to the Advisory Committee on Rules of Evidence: Rule Changes to Address Machine-Generated Proof Beyond Authentication,* Advisory Committee On Evidence Rules, Oct. 27, 2023, TAB 1E, pp. 80-81.

[8] There may be instances in which the creation of animations to demonstrate conditions not subject to ordinary observation can be quite useful. For example, in one of the patent infringement trials in which the author served as the government trial attorney at the U.S. International Trade Commission, the animation of flow dynamics within the semiconductor chip encapsulation process was extremely useful in determining whether the accused process practiced an asserted patent claim. *See In re Certain Dynamic Random Access Memories, Components Thereof, and Products Containing Same,* Inv. No. 337-TA-242, USITC Publication No. 2034 (Nov. 1987), https://www.usitc.gov/publications/337/pub2034.pdf.

[9] The Government Accountability Office defines "deepfake" as follows: "A deepfake is a video, photo, or audio recording that seems real but has been manipulated with AI. The underlying technology can replace faces, manipulate facial expressions, synthesize faces, and synthesize speech. Deepfakes can depict someone appearing to say or do something that they in fact never said or did." GAO, Science, Technology Assessment, and Analytics, SCIENCE & TECH SPOTLIGHT: DEEPFAKES, Feb., 2020, https://www.gao.gov/assets/gao-20-379sp.pdf.

[10] As discussed later in this article, the synthetic media section of the AI Action Plan also includes a reference to the TAKE IT DOWN Act, a recently enacted federal law that deals with non-consensual intimate visual depictions, such as "revenge porn," that might be genuine or created by AI deepfake technology. The full name of the Act is *Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks*, S. 146, 119th Cong. (2025), but is referenced herein by its common name, the "TAKE IT DOWN Act" or simply "the Act."

[11] *Kohn v. Ellison*, 2025 WL 66765 (D. Minn. 2025) *See also, e.g., Increasing Threat of Deepfake Identities*, Homeland Security, https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf.

[12] *Kohn v. Ellison*, 2025 WL 66765, p1. The *Kohn* case dealt with political speech – and the author of this article is indebted to Judge Provinzino for providing the following oasis of amusement in the midst of some of the graphic and very disturbing fact patterns in other deepfake cases: "False political speech is nothing new in our Nation. For example, in the 1800 presidential election, allies of Thomas Jefferson suggested that his rival, John Adams, was a woman. *See* Annie C. Hundley, *Fake News and the First Amendment: How False Political Speech Kills the Marketplace of Ideas*, 92 Tulane L. Rev. 497, 499–500 (2017). Adams's allies responded in kind by spreading a

The AI Action Plan initially phrases the problem as the need to ensure that "the courts and law enforcement" have the tools they need to overcome the new challenges posed by malicious deepfakes.[13] Our first thoughts likely turn to traditional courtrooms in which federal criminal and civil cases are adjudicated. However, there are numerous federal forums and processes – such as Federal Trade Commission proceedings, Securities and Exchange Commission proceedings, Commodity Futures Trading Commission proceedings, Federal Energy Regulatory Commission proceedings, U.S. Patent & Trademark Office proceedings, and U.S. International Trade Commission proceedings, to name a few – in which evidentiary standards are also a core consideration.[14] Outside of the realm of dispute resolution, evidence-based policymaking within the federal government also becomes a factor.[15] In short, the AI Action Plan's effort to address the threats posed by synthetic media will necessarily have broad applicability – well beyond the important context of typical judicial and law enforcement processes. That's OK. Deepfake technology affects us all, even if we are not in court or under indictment. The AI Action Plan recognizes this and seeks to expand its reach generally to all federal agencies that engage in adjudications.[16]

As the brief sampling above indicates, the potential proliferation of convincing deepfakes can snake through numerous channels that we have constructed for achieving a reasonable, just, and secure society. Of course, falsified evidence is nothing new – so, maybe we are in a clickbait fueled media world of the "sky is falling" variety. Maybe. However, common sense, shared notions of justice, and historical examples caution us to take this issue very seriously – as does the AI Action Plan. Many of us can imagine any number of worst-case scenarios, but history teaches us that some of these scenarios are very real. Can concocted evidence really corrupt entire systems as well as individual cases? Are we potentially that fallible? Are highly-positioned segments of society capable of utilizing lax or confounded evidentiary

---

rumor that Jefferson was dead. *Id.* This was news to Jefferson, who was very much alive (and went on to win the election)." 2025 WL 66765, p. 1.

[13] AI Action Plan, p. 13, https://whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

[14] For a good set of links to pertinent agency and government corporation sites, *see* United States Government Information Online: Federal Administrative Agencies, Pace University Elisabeth Haub School of Law, (updated Feb. 25, 2025), https://libraryguides.law.pace.edu/c.php?g=319332&p=2134043.

[15] *See, e.g., Evidence, Evaluation, and Learning, Foundations for Evidence-Based Policymaking Act of 2019*; U.S. Department of State, https://www.state.gov/evidence-evaluation-and-learning; Potok, *Data Usage Information and Connecting With Data Users: U.S. Mandates and Guidance for Government Agency Evidence Building*, Special Issue No. 4: Democratizing Data (April 2, 2024), As described in an excerpt from the article's Abstract the article delves into "[t]he key components of the Foundations of Evidence-Based Policymaking Act of 2018, the subsequent presidential executive orders and U.S. Office of Management and Budget memoranda aimed at promoting evidence-based practices, federal mandates for agencies to engage with data users, the value of interactions between agencies and data users, and the potential implications of those interactions for the future of evidence-based policy." https://hdsr.mitpress.mit.edu/pub/605i9bgg/release/3; *see also, Evidence Plans*, Evaluation/gov, https://www.evaluation.gov/evidence-plans/summary/.

[16] AI Action Plan, Second Recommended Policy Action, p. 13, https://whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

standards to weaponize the judicial system for gain, perversion, and unjust preservation of power? The following is an American "home grown" answer – and it points to George Santayana's cautionary statement: "Those who cannot remember the past are condemned to repeat it."[17]

II. Introductory Digression: Cotton Mather's Wonders Of The Invisible World Made Tangible – The Specter of Incredible Evidence Made Credible by Improper Motive, A Lack of Proper Scrutiny, And the Tendency to Believe That Which Is (Economically or Politically) Too Risky or Difficult to Question

Whether characterized as pertinent history or histrionic in this context, the "witch hunt" machinations employed in the early Massachusetts Bay Colony show the use and acceptance of falsified (or simply imaginary or hallucinated) evidence to strip the accused of their lives, liberties and properties – and to viciously enforce authority in a hierarchical political system. The strangeness of this situation is compounded by the fact that this was not an example of mob rule – it was a deliberate and methodically administered construction of a privileged authoritarian class, validating their political and social stations and enforcing their moral authority over those who (literally in one instance) were crushed under the weight of fake evidence. The following is a brief summary of the situation.

In 1692, Salem Village was embroiled in witch hunts that were premised, at least in part, on "spectral evidence." Generally defined, "spectral evidence" consists of third-party testimony that "the accused person's spirit or spectral shape appeared to the witness in a dream at the time the accused person's physical body was at another location."[18] As absurd as it now seems, one of the colony's brightest intellectual and theological luminaries – Harvard-educated Cotton Mather – actually endorsed the utility (*albeit* limited)[19] of spectral evidence. In his famous (or infamous) 1693 work, *The Wonders of the Invisible World,* Mather postulated that such evidence would be sufficient for an indictment although an actual conviction required more. Nonetheless, in light of the tortures and deprivations that could be inflicted on indicted persons as the hunt for more reliable evidence to support a conviction persisted, Mather's cautionary "half-way" acceptance of spectral evidence as a basis for indictment is now generally viewed as, at best, sadly misled,[20] and, at worst, a ruthless attempt to shore up a Puritan hierarchy and political base that was losing

[17] Santayana, *The Life of Reason: Reason in Common Sense,* Scribner's, 1905: 284; *see also*, Flamm, George Santayana, IEP, https://iep.utm.edu/santayan/.

[18] *State v. Dustin,* 122 N.H. 544, 551 (N.H. 1982) (disputed relevance of victim's apprehensions in criminal restraint case)(dissent).

[19] Mather's position on spectral evidence can seem a bit obscure, and sometimes seems to shift, but it is clear that he believed spectral evidence was sufficient to support an indictment but could not be used alone to support a conviction. Dorn, *Evidence From Invisible Worlds in Salem* (Aug. 20, 2020), https://blogs.loc.gov/law/2020/08/evidence-from-invisible-worlds-in-salem/.

[20] *See* comments of Marilynne Roach, *Cotton Mather And The Salem Witch Trials: Separating Fact From Fiction,* Aug. 8, 2024, Transcript, p. 11, https://salemwitchmuseum.com/videos/cotton-mather/.

its grip on the area's religious and social norms.[21] Also, recent scholarship indicates that the spectral raving of some key providers of evidence were likely scripted by Thomas Putnam, an ambitious cohort of Mather and a major influence of the Salem Witch Trials. In addition to capitalizing on "religious zeal," it appears that Putnam's secular motivations included political gain, aspiration to higher social status, or personal revenge – his signature being on the accusations of 14 of the 19 people executed under charges of witchcraft.[22]

In any case, with regard to unreliable "expert testimony" supporting concocted evidence and the resulting convictions, Cotton Mather's zeal went so far that he served as an expert to refute the notion that a practitioner of witchcraft was incapable of reciting The Lord's Prayer.[23] It happened this way: On August 19, 1692, ex-minister, religious non-conformist, debtor, and convicted witch George Burroughs was taken to Gallows Hill to be hanged, but when he flawlessly recited The Lord's Prayer several members of the crowd called for the execution to be halted. Mather intervened and, as an authority on the subject, explained to the crowd that Burroughs was convicted by a jury – and, besides, those possessed by the devil could mislead us by feigning piety.[24] The execution went forward – as did three others that day: John Proctor, John Willard, and Martha Carrier, neither of whom would, thanks to Mather, have the benefit of relying on a recitation of The Lord's Prayer.

So, did Mather pay a proper price or suffer any material social ostracism for his core role in the administrative and "evidentiary" side of the Salem executions? No – the opposite is true. Mather's continued political activism made him a key figure in the 1689 Boston Revolt that prefigured (at least in part) the spirit of self-determination that motivated the American

---

[21] Theories about the "outbreak" of persecution and witchcraft paranoia in Salem range from simple religious zeal, imported prejudices, trauma-induced mass hysteria, misogyny (ironically supported by women), class warfare, personal enmity, land grabs, feuding neighbors, ergot poisoning from rye (also called "St. Anthony's Fire), a surfacing vortex of spiritual degeneration, and other real or imagined inducers of behavioral abnormalities. However, while some of these factors may have been at work, the author's view is that no explanation is credible unless it tackles the ruthless acceptance of spectral evidence by a learned class of Puritan judges and aristocracy who administered the wave of terror with (relative) efficiency and deliberation – stopping only when the accusations became so bold as to include the governor's wife. *see*, Wallenfeldt, *Salem Witch Trials*, Britannica, (Updated July 31, 2025) where it is noted that the witch hunt mechanisms ground to a halt – and pardons were issued - after the wife of the colony's governor was accused of witchcraft. https://www.britannica.com/event/Salem-witch-trials/The-trials. Although beyond the scope of this article, contemporary examples of this "acceptable until personally threatening" phenomenon can be found almost daily by anyone who observes contemporary political and social developments. As noted elsewhere in this discussion, it sometimes seems that the more things change, the more they stay the same.

[22] Putnam himself accused and testified against 43 people, while his 12-year-old daughter, Ann Putnam, testified against 62 people. *See* Brooks, Thomas Putnam: *Ringleader of the Salem Witch Hunt?,* History of Massachusetts Blog, Nov. 19, 2013, https://historyofmassachusetts.org/thomas-putnam-ringleader-of-the-salem-witch-hunt/. *See also,* Wemble, *The Putnams And The Salem Witch Trials*, Wingrave-with-Rowsham Heritage Association, https://wingrave-rowsham-heritage.org.uk/articles/the-putnams-and-the-salem-witch-trials/.

[23] Linder, *Famous Trials, Accounts and Materials for 100 of History's Most Important Trials, Cotton Mather*, MKC School of Law, https://famous-trials.com/salem/2037-sal-bmat.

[24] Brooks, *The Witchcraft Trial of Reverend George Burroughs*, History of Massachusetts Blog, April 9, 2017, https://historyofmassachusetts.org/reverend-george-burroughs-salem/.

Revolution.[25] He also conducted experiments in plant hybridization and he was elected a member of the Royal Society of London in 1713, due in part to his work on inoculation as a means to combat smallpox. In light of his frequent devotion to the empirical process, it is difficult to imagine that his endorsement of spectral evidence stemmed from anything other than his dedication to advancing the interests of the Puritan hierarchy and its willingness to believe evidence that supported its religion-based power and social dominance. Perhaps Mather was sincere, but he accepted imaginary evidence to support ruthless prosecutions for imaginary crimes. Therefore, we are left with the question: How could such public-spirited advocates and learned judges be misled on fundamental questions of reliable and credible evidence? The best (*albei*t not fully adequate) answer seems to be that this is one of those "red flags" that history has passed down from generation to generation – and we need to keep it in mind: Judges and juries are only as good as the rules they choose to follow, the quality of the evidence made available to them, and the resources available to properly evaluate, authenticate or discredit that evidence.

In *Wonders*, Mather advanced a number of convoluted religious, philosophical, and truth-seeking motives to justify preliminary reliance on spectral evidence. He did this in the context of an oppressive wave that saw almost ten percent of the population in and around Salem Village accused of the crime of witchcraft within a roughly one-year period – with nineteen executions by hanging, one victim "pressed" to death by heavy stones during an attempt to elicit a confession, and five other deaths caused by conditions of incarceration.[26] This is terrible stuff. So, where do AI deepfakes come in?

We now have the widespread ability to readily create "tangible" AI-assisted fake evidence that Mather and his contemporaries could only imagine.

Consequently, we need to provide our people and forums with ready access to the technological assistance necessary to combat injurious AI-assisted deepfakes so there is a solid basis for the application of the relevant Rules of Evidence – which are, after all, only as good as the resources available for their proper implementation.

III. Adjusting To The Wonders Of The Deepfake World

[25] *See* Oliver (Editor) *The Great Boston Revolt Of 1689*, New England Historical Society (updated 2023), https://newenglandhistoricalsociety.com/great-boston-revolt-1689/. The newly ascended joint monarchs, William and Mary, who were Protestants, supported the revolt against officials appointed by the former king, James II, a Catholic and generally (although not always) a supporter of religious tolerance. It was James II's attempt to reorganize the government of the colony, appoint new officials, and loosen the Puritan grip on the population that led to hostility on the part of local Puritan elites, including Mather. *Id.*; Brooks, *How Did the Glorious Revolution in England Affect the Colonies?* History of Massachusetts Blog, Jan. 19, 2016, https://historyofmassachusetts.org/how-did-glorious-revolution-affect-colonies/; Cotton Mather Biography, https://www.poemhunter.com/cotton-mather/.

[26] Scanlan, *The Salem Witch Trials According to the Historical Records*, HUMANITIES, Winter 2022, Vol. 43, No. 1, https://www.neh.gov/article/records-salem-witch-trials; *see also*, Wallenfeldt, *Salem Witch Trials*, Britannica, (Updated July 31, 2025). https://www.britannica.com/event/Salem-witch-trials/The-trials.

A. Introduction To A Coherent And Active Executive Initiative (Working With The Judicial Branch)

The AI Action Plan notes that AI-generated media may present novel challenges to the legal system – the proposal being that "[t]he Administration must give the courts and law enforcement the tools they need to overcome these new challenges."[27] To facilitate this process, the following three Recommended Policy Actions are provided in the Action Plan:

> 1. Led by the National Institute of Standards and Technology (NIST) at the Department of Commerce (DOC), consider developing NIST's *Guardians of Forensic Evidence* deepfake evaluation program into a formal guideline and a companion voluntary forensic benchmark.
>
> 2. Led by the Department of Justice (DOJ), issue guidance to agencies that engage in adjudications to explore adopting a deepfake standard similar to the proposed Federal Rules of Evidence Rule 901(c) under consideration by the Advisory Committee on Evidence Rules.
>
> 3. Led by DOJ's Office of Legal Policy, file formal comments on any proposed deepfake- related additions to the Federal Rules of Evidence.[28]

The First Recommended Policy Action's reference to NIST's *Guardians of Forensic Evidence* deepfake evaluation program cites a set of NIST presentation slides titled *Guardians of Forensic Evidence: Evaluating Analytic Systems Against AI-Generated Deepfakes*.[29] The NIST presentation emphasizes that the development and deployment of Advanced Analytic Tools "are critical to preserving the integrity of forensic evidence and ensuring justice in the digital age."[30] This recognition is premised on notions that AI-generated deepfakes pose significant challenges in "legal frameworks," as: (1) courts and law enforcement are struggling to keep pace with the rapid growth of AI-powered fraud; (2) AI-generated disinformation erodes public trust in media platforms and complicates forensic investigations; and (3) deepfake assisted fraud or

---

[27] *See* Pillar 1: Accelerate AI Innovation, *Combat Synthetic Media in the Legal System*, AI Action Plan, pp. 12-13, https://whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf

[28] *Id.* at p. 13. The reference in the first Recommendation to NIST's *Guardians of Forensic Evidence* deepfake evaluation program specifically notes the following source: Haiying Guan, James Horan, and Andrew Zhang, *Guardians of Forensic Evidence: Evaluating Analytic Systems Against AI-Generated Deepfakes,* (Gaithersburg, MD: National Institute of Standards and Technology, January 27, 2025, www.nist.gov/publications/guardians-forensic-evidence-evaluating-analytic-systems-against-ai-generated-deepfakes .

[29] Guan, Zhang, and Horan, Slide Set, *Guardians of Forensic Evidence: Evaluating Analytic Systems Against AI-Generated Deepfakes,* Multimodal Information Group (MIG), Information Access Division (IAD), Information Technology Laboratory (ITL), National Institute of Standards and Technology (NIST), Published Jan. 27, 2025, https://www.nist.gov/publications/guardians-forensic-evidence-evaluating-analytic-systems-against-ai-generated-deepfakes.

[30] *Id.* at Slide 2.

misinformation causes massive economic loss across a range of industries, including, for example, in financial services and fintech industries.[31]

The Program Goals envisioned in NIST's *Guardians of Forensic Evidence* are:

1. Foster research in Deepfake and Generative AI;
2. Conduct recurring evaluations for state-of-the-art insights;
3. Collaborate with academia/industry to establish a reference baseline detection system;
4. Provide performance analysis for iterative system improvement;
5. Support the transition from lab prototypes to real-world products;
6. Enhance generalization of detection tools; and
7. Deliver cross-year comparison reports.[32]

There are numerous technologies available and developing to detect deepfakes. However, as technologies and techniques develop, AI developers are quick to generate "improvements" and counter-measures. The old "folk technique" of looking at the rendering of hands in likely AI-generated art – such as the cover of the Willie Crowder recording, *She Left Me For A Man With Shoes*[33] – to detect the formerly notorious inability of AI to generate convincing depictions of human hands doesn't work in the case of more developed AI systems. This rapid outmoding is also present in more sophisticated techniques. For example, Intel developed an AI-driven "FakeCatcher" system that uses (get ready to pronounce this) photoplethysmography,[34] which detects changes in blood flow that affects subtle color changes in human faces with every heartbeat – changes that were not exhibited by the deepfake technology of the time – but a criticism is that the system is not completely accurate, depends on source image quality, and can become outmoded without notice as deepfake-creation AI becomes more sophisticated. As one commentator put it, "[a]lthough much is possible, the prospect of ever preventing all dangerous deepfakes, which are growing more nuanced all the time, seems like it might be beyond even tireless AI."[35] It is important that we pay attention to the skeptical voices among us because they

---

[31] *Id.*
[32] *Id.* at Slide 4.
[33] *See, She Left Me for a Man With Shoes* – Willie Crowder (1933 Lost Delta Blues) Dumpster Groves, https://www.youtube.com/watch?v=BzNygRNNA1E&list=RDBzNygRNNA1E&start_radio=1;
See also, Your Mama Don't Like Me (And She Might Be Right) - Elijah "Hollowfoot" Turner (Rare 1937 Delta Blues), Dumpster Groves, https://www.youtube.com/watch?v=gI7zgDmjiyY.
[34] Photoplethysmography is a non-invasive method for measuring blood volume changes in a microvascular bed of the skin based on optical properties, such as absorption, scattering, and transmission properties of human body composition under a specific light wavelength. *See, e.g.,* Park, Seok, Kim & Shin, *Photoplethysmogram Analysis and Applications: An Integrative Review*, Frontiers in Physiology, Vol. 12 -2021, https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2021.808451/full.
[35] Nash, *Intel shows its FakeCatcher but deepfake's challenge might be too big*, Biometric Update. Com, (July 25, 2023)
https://www.biometricupdate.com/202307/intel-shows-its-fakecatcher-but-deepfakes-challenge-might-be-too-big.

help to identify flaws in technologies and perceptions – and, therefore, render valuable assistance in our drive to meaningfully combat deepfakes. However, our understanding that there is no perfect solution to an ever-evolving problem doesn't mean we can afford to simply leave it unattended.

The key is that there needs to be a well-funded, generally-available, continually updated, and objective set of technological resources to assist in the identification and discrediting of deepfake evidence. This approach posits a publicly-funded "expert" system that would be available throughout the range of evidence-driven processes. Still, do we need to be careful about ceding too much authority to one expert, agency or organization? The answer is definitely "yes," and here's an example that supports that answer: Perhaps one of the more bizarre deepfake-related cases, *Kohls v. Ellison*, involved the exclusion of an expert declaration on AI, deepfakes, and the dangers of deepfakes to free speech and democracy. [36] The problem was that the declarant, "a credentialed expert on the dangers of AI and misinformation" relied on AI-hallucinated content in his declaration, cited two non-existent academic articles, and incorrectly cited the authors of a third article. As the Court put it, "a credentialed expert on the dangers of AI and misinformation, has fallen victim to the siren call of relying too heavily on AI—in a case that revolves around the dangers of AI, no less."[37] Whether your source is the Roman satirist Juvenal or *Star Trek: The Next Generation*, the formulation of the problem persists in one form or another: "Who is to stand guard over the guards themselves?" [38] or "Who Watches the Watchers?" [39] Again, think of Cotton Mather's unchecked rendering of biased expertise. However, again, the unending nature of this problem is not a basis for giving up. As in a number of generative AI systems,[40] the use of an adversarial process to achieve better results immediately suggests itself. NIST's *Guardians of Forensic Evidence* deepfake evaluation program is a very welcome development, but we will always need additional competitive, critical organizations – perhaps at the State level, in the not-for-profit sphere, and in international cooperative contexts – to ensure that efforts such as the NIST's program remain on track, provide reliable guidance, and alternative voices are heard and considered in maintaining the system.

The Second and Third Recommended Policy Actions in the synthetic media section of the AI Action Plan seek to reinforce or improve evidentiary standards so they can more readily identify

---

[36] 2025 WL 66514 (D. Minn. 2025).
[37] *Id.* at p. 4.
[38] Juvenal, *Satire VI*, *The Sixteen Satires*, Translated by Peter Green, Penguin Books, Third Edition, 2004, p. 45, lines 31-32; *See, also, e.g., Quis custodiet ipsos custodes?* Hull AWE, https://hull-awe.org.uk/index.php/Quis_custodiet_ipsos_custodes%3F; Jonas, *CRIMINAL JUSTICE, Who's guarding the guards? State police scandal only getting worse,* CommonWealth Beacon, Politics, Ideas, and Civic Life in Massachusetts (Oct. 11, 2018), https://commonwealthbeacon.org/criminal-justice/whos-guarding-the-guards-2/.
[39] *Star Trek: The Next Generation*, Episode 4, Third Season, *Who Watches the Watchers*, written by Richard Manning and Hans Beimier; first aired Oct. 16, 1989. For a discussion of the episode and details of its casting and production, *see Who Watches The Watchers (episode)*, Fandom, Memory Alpha, https://memory-alpha.fandom.com/wiki/Who_Watches_The_Watchers_(episode).
[40] *See, e.g.,* Stryker and Scapicchio*, What is generative AI? IBM*, (March 22, 2024), *https://www.ibm.com/think/topics/generative-ai; What is a GAN*, https://aws.amazon.com/what-is/gan/.

and discredit deepfakes. As noted above, the Second Recommended Policy Recommendation specifically references proposed section (c) to Federal Rule of Evidence 901.[41] However, there are other Rules that come into play, such as: (1) Federal Rule of Evidence 104, which empowers the judge in a case to rule on preliminary questions of admissibility,[42] (2) Federal Rule of Evidence 902 that deals with evidence that is self-authenticating,[43] and (3) Federal Rule of Evidence 706, which empowers judges to appoint expert witnesses who assist the court in understanding complex technical or specialized matters.[44] In short, passing the threshold of Rule 901 (or 902) keeps you in the race, but it doesn't guarantee you'll get past the hurdle of admissibility – and, therefore, the AI Action Plan is not limited to Rule 901.

As one commentator put it, "Rule 901 basically operates to prevent the jury from wasting its time evaluating an item of evidence that clearly is not what the proponent claims it to be."[45] A determination of authenticity can then place the evidence in a condition for assessment as to admissibility.[46] However, many would agree that Rule 901 poses a pretty low threshold for establishing authenticity.[47]

The text of the current form of Rule 901 is as follows:

> (a) In General. To satisfy the requirement of authenticating or identifying an item of evidence, the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is.
>
> (b) Examples. [1 through 9][48]

---

[41] 28 USC App, FEDERAL RULES OF EVIDENCE, ARTICLE IX: AUTHENTICATION AND IDENTIFICATION, From Title 28—Appendix, FEDERAL RULES OF EVIDENCE, https://uscode.house.gov/view.xhtml?path=/prelim@title28/title28a/node230/article9&edition=prelim.

[42] *See Rule 104. Preliminary Questions*, Legal Information Institute, Cornell Law School, https://www.law.cornell.edu/rules/fre/rule_104.

[43] The conditions of self-authentication should be considered with regard to deepfake technology's ability to mimic at least some of the conditions of self-authentication. For example, as noted in Advisory Committee On Evidence Rules, May 2, 2025, Tab 3, Part IX Draft of a Deepfake Amendment, p. 69: "To take the most obvious example, newspapers and periodicals, which are self-authenticating under Rule 902(7) can definitely be deepfaked." https://www.uscourts.gov/sites/default/files/document/2025-05_evidence_rules_committee_agenda_book_final.pdf.

[44] *See, e.g.,* Barreto, *Federal Rule 706: Court-Appointed Experts for Impartial Insight*, Expert Institute, U[dated March 5,2025, https://www.expertinstitute.com/resources/insights/federal-rules-of-evidence-706/.

[45] Capra, *"Deepfakes" and Possible Amendments to Article 9 of the FRE*," Advisory Committee On Evidence Rules, Oct. 27, 2023, TAB 2A, p.87.

[46] *See, e.g.,* Federal Rule of Evidence 104(a). *See also, United States v. Whitehead*, 2024 WL 3085019*9 (S.D.N.Y. 2024)(focusing on chain of custody); *citing United States v. Sovie*, 122 F.3d 122, 127 (2d Cir. 1997); *United States v. Reid*, 650 F. Supp. 3d 182, 196 (S.D.N.Y. 2023); *and United States v. Tropeano*, 252 F.3d 653, 661 (2d Cir. 2001).

[47] *See, e.g.,* Advisory Committee On Evidence Rules, May 2, 2025, *What is the Evidentiary Problem Raised by Machine Learning, III. Basic Rules on Authenticity*, p. 173, https://www.uscourts.gov/sites/default/files/document/2025-05_evidence_rules_committee_agenda_book_final.pdf.

[48] The Examples section of the Rule provides nine examples and reads as follows: "The following are examples only—not a complete list—of evidence that satisfies the requirement:(1) *Testimony of a Witness with Knowledge.* Testimony that an item is what it is claimed to be. (2) *Nonexpert Opinion About Handwriting.* A nonexpert's opinion that handwriting is genuine, based on a familiarity with it that was not acquired for the current litigation. (3) *Comparison by an Expert Witness or the Trier of Fact.* A comparison with an authenticated specimen by an expert witness or the trier of fact. (4) *Distinctive Characteristics and the Like.* The appearance, contents,

A proposed new Section (c) was recently considered by the Advisory Committee On Evidence Rules[49] The text of the proposed Section (c) is as follows:

> Rule 901(c). Potentially Fabricated Evidence Created By Generative Artificial Intelligence.
>
> If a party challenging the authenticity of an item of evidence demonstrates to the court that a jury reasonably could find that the item has been fabricated, in whole or in part, by generative artificial intelligence, the item is admissible only if the proponent demonstrates to the court that it is more likely than not authentic.
>
> This rule governs authentication under both Rule 901 and 902.[50]

Another proposed revision, offered by Professor Rebecca A. Delfino,[51] is as follows:

> Notwithstanding subdivision (a), if a party challenging the authenticity of computer-generated or other electronic evidence presents evidence sufficient to support a factual finding that the challenged evidence has been manipulated or fabricated, in whole or in part by generative artificial intelligence, the proponent of the evidence must authenticate the evidence under subdivision (b) and provide additional proof

---

substance, internal patterns, or other distinctive characteristics of the item, taken together with all the circumstances.(5) *Opinion About a Voice.* An opinion identifying a person's voice—whether heard firsthand or through mechanical or electronic transmission or recording—based on hearing the voice at any time under circumstances that connect it with the alleged speaker. (6) *Evidence About a Telephone Conversation.* For a telephone conversation, evidence that a call was made to the number assigned at the time to: (A) a particular person, if circumstances, including self-identification, show that the person answering was the one called; or (B) a particular business, if the call was made to a business and the call related to business reasonably transacted over the telephone. (7) *Evidence About Public Records.* Evidence that: (A) a document was recorded or filed in a public office as authorized by law; or (B) a purported public record or statement is from the office where items of this kind are kept. (8) *Evidence About Ancient Documents or Data Compilations.* For a document or data compilation, evidence that it: (A) is in a condition that creates no suspicion about its authenticity; (B) was in a place where, if authentic, it would likely be; and (C) is at least 20 years old when offered. (9) *Evidence About a Process or System.* Evidence describing a process or system and showing that it produces an accurate result.(10) *Methods Provided by a Statute or Rule.* Any method of authentication or identification allowed by a federal statute or a rule prescribed by the Supreme Court."

[49] The Supreme Court first established a rules advisory committee in June of 1935 to help draft the Federal Rules of Civil Procedure, which took effect in 1938. Today, Advisory Committees on the Rules of Appellate, Bankruptcy, Civil, Criminal Procedure, and the Rules of Evidence carry on a continuous study of the rules and recommend changes to the Judicial Conference through a Standing Committee on Rules of Practice and Procedure. The Chief Justice appoints the committee members whose terms are limited to no more than six years. Committee members receive no payment for their service. Unlike other Judicial Conference committees, the rules committees include not only federal judges, but also practicing lawyers, law professors, state chief justices, and high-level officials from the Department of Justice and federal public defender organizations. *See* United States Courts, *Committee Membership Selection,*
https://www.uscourts.gov/forms-rules/about-rulemaking-process/committee-membership-selection#:~:text=Today%2C%20Advisory%20Committees%20on%20the,documenting%20the%20rules%20committees'%20work.

[50] Advisory Committee On Evidence Rules, May 2, 2025, Tab 3, Part IX Draft of a Deepfake Amendment, p. 69, https://www.uscourts.gov/sites/default/files/document/2025-05_evidence_rules_committee_agenda_book_final.pdf.

[51] Professor Delfino's CV (current as of Aug. 18, 2025) notes that she is an Associate Professor of Law (Research Tenure-Track) (2025-Present), LMU Loyola Law School, Los Angeles.
https://www.lls.edu/faculty/facultylistc-d/rebeccadelfino/.

> establishing its reliability. The court must decide the admissibility of the challenged evidence under Rule 104(a).[52]

The language of each proposed new Section (c) keeps the burden of credibly challenging alleged deepfake evidence on the party opposing the evidence. This is in keeping with the general notion that it is not the proponent's duty to prove that the proffered item is not a deepfake - the burden is on the opponent to show the likelihood of a deepfake.[53] This position is decisive and clean, except for those who can't afford to engage expert witnesses or technologies to help assess the potentially AI-altered or wholly AI-generated evidence. Should we ever have to revert to the questions: Can you afford justice? Does the average litigant, accused, or advocate in a fact-finding process have to rely on the sometimes-elusive availability of suitable experts? How do we ensure that the providers of verification technologies are objective and the technologies are not weaponized along personal, ideological or political lines? Are we back to 1692 Salem Village, confronting concocted evidence "without a prayer"?

Professor Rebecca A. Delfino, the author of a proposed new Rule 901(c), offered the following thoughts on experts and the cost of combating deepfakes.

> In cases involving deepfakes, the courts should use their *sua sponte* powers under Federal Rule of Evidence 706 to appoint independent expert witnesses to assist the court in understanding the deepfake evidence and allegations. In addition, in all deepfake cases, including those where the parties seek to retain their own digital forensic experts, the proponent of the deepfake allegation should bear the cost of proving it unless the court determines, based on financial need, that costs to litigate non-frivolous deepfake evidence claims should be allocated to the other party.[54]

There is excellent reasoning and sensitivity in Professor Delfino's proposal – and her approach may ultimately be routinely adopted. However, determining the threshold issue of whether a case involves a deepfake might simply be too much without the type of assistance of a program of the type envisioned or hinted at in the AI Section Plan's First Recommended Policy Action.

---

[52] Advisory Committee On Evidence Rules, May 2, 2025, TAB 3, Part VI., pp. 53-54, https://www.uscourts.gov/sites/default/files/document/2025-05_evidence_rules_committee_agenda_book_final.pdf; Rule 104(a) relates to preliminary questions of admissibility generally: "**(a) In General.** The court must decide any preliminary question about whether a witness is qualified, a privilege exists, or evidence is admissible. In so deciding, the court is not bound by evidence rules, except those on privilege." *See* Cornell Law School, LII Legal Information Institute, https://www.law.cornell.edu/rules/fre/rule_104.

[53] *See, e.g.,* Advisory Committee On Evidence Rules, May 2, 2025, TAB 3, p. 39, https://www.uscourts.gov/sites/default/files/document/2025-05_evidence_rules_committee_agenda_book_final.pdf; *see also Matter of Gabriel H.,* 229 A.D.3d 1048, 215 N.Y.S.3d 613 (4th Dept. (2024); *Pittman v. Commonwealth*, No. 0681-22-1, 2023 WL 3061782, at *6–7 (Va. Ct. App. Apr. 25, 2023). Note, however, variations can occur depending on the type of evidence under consideration. For example, the Second Circuit has "adopted a general standard, namely, that the [proponent] produce clear and convincing evidence of authenticity and accuracy as a foundation for the admission of" recorded evidence. *US v. Whitehead*, 2024 WL 3085019*9 (S.D.N.Y. 2024) *citing United States v. Ruggiero*, 928 F 2d 1289, 1303 (2d Cir. 1991).

[54] *Id.* at p. 3, *citing* Rebecca Delfino, *Pay-to-Play: Access to Justice in the Era of AI and Deepfakes,* 55 Seton Hall L. Rev. 789 (2025), Loyola Law School, Los Angeles Legal Studies Research Paper No. 2024-08, Available at SSRN: https://ssrn.com/abstract=4722364 or http://dx.doi.org/10.2139/ssrn.4722364.

Put another way, it's time to recognize that the public, the judicial system, and other evidence-driven processes need a generally available, objective, and continually updated technological resource to detect and weed out deepfakes – perhaps with a very vigorous enforcement of criminal laws against presenting false evidence.[55] Also, it may be time to consider that critical evidence subject to deepfake creation or manipulation – such as video of an alleged crime being committed – should require independent expert certification secured by the proponent of the evidence. Notably, we have a federal resource for fingerprint and criminal history data along with modest charges for, *e.g.,* use of such federal resources in civil matters.[56] There are also resources, such as the State Justice Institute, that provide grants and works with courts to address a range of challenges, including AI-related challenges.[57] Also, the Innocence Project,[58] a publicly funded organization, combats, *e.g.,* the misapplication of forensic science in prosecuting innocent individuals in our criminal system.[59] In short, there are a number of models that can provide examples – or inspiration – that can help to shape a broader and more readily available resource for preventing the use of improper deepfakes in evidence-driven proceedings. After all, no matter how "perfect" current or amended evidentiary rules might be, they are fundamentally hollow if there are no readily available resources to ensure that they can be implemented in a just and practical way. That is why the AI Action Plan's First Recommended Policy Action is so critical; it points toward the development of an organization and process for

---

[55] *See*, *e.g.,* Doyle, *False Statements and Perjury: An Overview of Federal Criminal Law*, CRS Products (Library of Congress), 10/08/2024, https://www.congress.gov/crs-product/98-808. As Doyle points out: "Federal criminal law features four general statutes that proscribe providing false information in matters relating to the federal government. One statute, 18 U.S.C. § 1001, proscribes false statements in matters within the jurisdiction of a federal agency or department. A second, 18 U.S.C. § 1621, condemns perjury with respect to any matter in federal law given under oath or penalty of perjury. The third, 18 U.S.C. § 1623, outlaws false declarations before federal grand juries or courts. The fourth, 18 U.S.C. § 1622, criminalizes inducing another to commit a federal perjury offense. Finally, conspiracy to commit any of these underlying crimes is a separate federal crime. Moreover, a defendant under investigation or on trial for some other federal offense may find upon conviction his sentence for the underlying offense enhanced as a consequence of a false statement made during the course of the investigation or trial." (citations omitted.). *Id.* at p. 1.

[56] The Integrated Automated Fingerprint Identification System, more commonly known as the IAFIS, is a national fingerprint and criminal history system maintained by the FBI's Criminal Justice Information Services (CJIS) Division. The IAFIS provides automated fingerprint search capabilities, latent search capabilities, electronic image storage, and electronic exchange of fingerprints and responses, 24 hours a day, 365 days a year. The electronic submission of fingerprints allows agencies to receive electronic responses to criminal tenprint fingerprint submissions within two hours and within 24 hours for civil fingerprint submissions. *See* IAFIS, The Integrated Automated Fingerprint Identification System, U.S. Department of Justice, Federal Bureau of Investigation, Criminal Justice Information Services Division, https://ucr.fbi.gov/fingerprints_biometrics/biometric-center-of-excellence/files/iafis_0808_one-pager825. There is generally a modest charge for use of federal resources of the kind discussed in, *e.g.,* civil litigation or investigative contexts. *See, e.g.,* FBI Criminal Justice Information Services Division; User Fee Schedule, Aug. 14, 2024, https://www.federalregister.gov/documents/2024/08/29/2024-19086/fbi-criminal-justice-information-services-division-user-fee-schedule.

[57] *See, Supporting the Nation's Judicial System & the Public it Serves,* State Justice Institute, https://www.sji.gov/priority-investment-areas/technology/.

[58] *See* About, the Innocence Project, https://innocenceproject.org/about/.

[59] *See Misapplied forensic science contributed to more than half of our wrongful conviction cases and nearly a quarter of all wrongful conviction cases since 1989*, the Innocence Project, https://innocenceproject.org/misapplication-of-forensic-science/.

continually battling and preventing the intrusion of damaging deepfakes into the system – sort of like the security program that is constantly updated to protect your computer.[60]

The Third Policy Recommendation in the AI Action Plan – "Led by DOJ's Office of Legal Policy, file formal comments on any proposed deepfake-related additions to the Federal Rules of Evidence" – simply reinforces the notion that federal resources need to be marshalled and dedicated to providing the expertise necessary to help shape any amendments that are intended to address deepfakes. As previously noted, while the Plan specifically mentions Federal Rule of Evidence 901, there are a number of areas where rule changes or creation come into play. For example, on June 10, 2025, the Judicial Conference Committee on Rules of Practice and Procedure (Standing Committee) approved publication of proposed new Rule 707.[61] The Committee prefaced its publication, comments, and call for public comments by noting:

> At its Fall meeting, the Committee considered proposals to amend the Evidence Rules to regulate machine learning and deepfakes. As to machine learning, the concern is that it might be unreliable, and yet the unreliability will be buried in the program and difficult to detect. The hearsay rule is likely to be inapplicable because the solution to hearsay is cross-examination, and a machine cannot be cross-examined. The Committee determined that the reliability issues attendant to machine output are akin to those raised by experts under Rule 702. Indeed, Rule 702 would be applicable to machine-learning if it was used by a testifying expert to reach her conclusion. But Rule 702 is not clearly applicable if the machine output is admitted without any expert testimony – either directly or by way of a lay witness. [62]

The text of proposed new Rule 707 is as follows:

> Rule 707. Machine-Generated Evidence[63]
>
> When machine-generated evidence is offered without an expert witness and would be subject to Rule 702 if testified to by a witness, the court may admit the evidence only if it satisfies the requirements of Rule 702(a)-(d). This rule does not apply to the output of simple scientific instruments.[64]

---

[60] The details of such an Executive Branch "support system" for evidence-driven forums (and whether it should originate and be shaped via legislation) will need to be considered in a future discussion. One concern, likely of many, is that the system should not be a routine source of unnecessary delay or "gamesmanship" by participants in the proceedings.

[61] Preliminary Draft, Proposed Amendments Published for Public Comment, Part V, pp. 100-111, https://www.uscourts.gov/forms-rules/proposed-amendments-published-public-comment. The comment period specified is from August 15, 2025 to February 16, 2026.

[62] *Id.* at p. 102.

[63] *Id.* at p. 109.

[64] As explained by the Committee: "The rule provides that it does not apply to the output of basic scientific instruments, and the Committee Note provides examples of such instruments, such as a mercury-based thermometer, an electronic scale, or a battery-operated digital thermometer. The Committee concluded that such an exception is warranted to avoid litigation over the output of instruments that can be presumed reliable but that, given the wide range of potential instruments and technological change, it is better to leave it to judges to determine whether a

Rule 702(a)-(d) relates to testimony by expert witnesses and provides:

> A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if the proponent demonstrates to the court that it is more likely than not that:
>
> **(a)** the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
>
> **(b)** the testimony is based on sufficient facts or data;
>
> **(c)** the testimony is the product of reliable principles and methods; and
>
> **(d)** the expert's opinion reflects a reliable application of the principles and methods to the facts of the case.[65]

The tensions and predicaments are clear. For example, where are the courts supposed to obtain sufficient expertise to apply proposed Rule 707 – and satisfy Rule 702(a)-(d) – unless they revert to securing expert testimony under Federal Rule of Evidence 706? Again, a generally available public resource suggests itself – likely developed with involvement of NIST's *Guardians of Forensic Evidence* deepfake evaluation program referenced in the AI Action Plan's pertinent Recommended Policy Action.[66] Publication of the proposed Rule 707 for public comment is not necessarily an endorsement by the Committee; the proposal is released simply to gather input.[67] The key, however, is that the Committee recognizes the gravity of the situation and is attempting to address it.

Before we leave this topic of deepfakes and evidence, it is noteworthy that, as public awareness of deepfake technology and its capabilities increases, skepticism increases and produces what has been termed "The Liar's Dividend."[68] One aspect of The Liar's Dividend is that the public –

---

particular instrument falls within the exception than to try to be more specific in the rule. The Committee Note also provides that the rule not apply to output that can be judicially noticed as reliable." *Id.* at p. 102-103.

[65] *See, e.g.,* Cornell Law School, LII Legal Information Institute, Rule 702. Testimony by Expert Witnesses, https://www.law.cornell.edu/rules/fre/rule_702; *see also, Federal Rule of Evidence 702, Judicia Conference Amends Rule 702*; 138 Harv. L. Rev. 899, Jan. 2025, https://harvardlawreview.org/print/vol-138/federal-rule-of-evidence-702/.

[66] *See* Pillar 1: Accelerate AI Innovation, *Combat Synthetic Media in the Legal System*, AI Action Plan, pp. 12-13, https://whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

[67] At its meeting, the Committee, by a vote of 8-1, recommended the proposal to add a new Rule 707 for release for public comment. The Department of Justice (DOJ) voted against the proposal. *Id.* at p. 103. The DOJ representative took the position that Rule 702 already covers the use of machine-generated evidence, and that Rule 707 only seeks to predict and regulate future needs. *See* Hill, *Proposed Rule 707 Targets AI-Crafted Evidence*, The National Law Review, Vol. XV, No. 234, June 5, 2025, https://natlawreview.com/article/proposed-rule-707-targets-ai-crafted-evidence#:~:text=Indeed%2C%20the%20point%20of%20the,and%20critique%20the%20proposed%20rule.

[68] Chesney & Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, California Law Review, Vol. 107, pp. 1785-86, Dec. 2019 (capitalization supplied), https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security, I have taken the liberty of using initial letter capitalization in "The Liar's Dividend" to indicate that it is a

including juries, judges, and other factfinders – will become overcautious in accepting or assigning weight to what would otherwise be persuasive evidence. As one set of commentators put it: "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."[69] While those commentators move their thesis into political bias, the general proposition is sound and further illustrates how deepfakes can – unless susceptible to readily available and reliable counter-technologies – jump from the confines of the Rules of Evidence into more general "psychological" areas of the deliberative process. For example, a somewhat analogous situation arises when the public becomes familiar with dramas or documentary-type depictions of criminal cases that involved relatively sophisticated evidence, such as cutting-edge DNA evidence, or sophisticated evidence gathering techniques – the result being that the bar of "beyond a reasonable doubt" is sometimes improperly raised because jurors misunderstand or are skeptical of cases that do not present sophisticated evidence or cutting-edge evidence-gathering techniques. This effect, sometimes called "the CSI effect," can lead to improper acquittals due to unwarranted juror expectations – with one juror being overheard to complain that the prosecution did not do a thorough job because "they didn't even dust the lawn for fingerprints."[70] Think of it this way: beyond the barriers presented by the Rules of Evidence and the forums' role in weeding out deepfake evidence there is potentially another obstacle presented by the (human) factfinder's AI-conscious mindset and predispositions – and, for example, limiting our consideration to rules of evidence that operate outside of the jurors' box would be a serious mistake. The AI Action Plan's First Policy Recommendation – the establishment of a proper deepfake vetting system – can also help to mitigate The Liar's Dividend. It's a matter of education initiatives and well-founded confidence in the system.

Part 3: Removal Procedures And Criminal Penalties For Nonconsensual Intimate Visual Depictions: An Introduction To The TAKE IT DOWN Act, Its Applications, Mandatory Notice And Takedown Provisions And Criminal Sanctions; With Ancillary Musings On Nonconsensual Intimate Audio Works And Allegedly Injurious Large Language Models – Should We Be Talking To A Libertine Chatbot About This?

The machine's danger to society is not from the machine itself but from what man makes of it. – Norbert Wiener[71]

---

specific referent to a species of unearned advantage in the context of deepfakes and generative AI. This is simply a convention to distinguish it from any number of political or social observations about the consequences (or perceived advantages) of deceit.

[69] *Id.* at 1785.

[70] Shelton, *The 'CSI Effect": Does It Really Exist?,* National Institute of Justice Journal, March 16, 2008, https://nij.ojp.gov/topics/articles/csi-effect-does-it-really-exist.

[71] *See From Cybernetics to AI: the pioneering work of Norbert Wiener*, MP Neuro, Max Planck Neuroscience, Max Planck Institute For Biological Cybernetics, April 25, 2024, https://maxplanckneuroscience.org/from-cybernetics-to-ai-the-pioneering-work-of-norbert-wiener/. An excellent insight into Wiener's broad vision, rigorous thought, and talent for anticipation – and the context for the quote – can

I. A Brief History of ~~Time~~ Skullduggery

As a means to set the context for the *Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act – i.e.,* the TAKE IT DOWN Act[72] – (hereinafter "the TAKE IT DOWN Act" or "the Act") some preliminary thoughts and more serious observations about deepfake audio works and, also, a particularly damaging Large Language Model AI-driven trend that needs to be urgently addressed, are provided.

The unfortunate human characteristic of deception and fakery seems to have been ever-present, even so much as to be (very forcefully) addressed in the First Law of The Code of Hammurabi.[73] So, the old adage persists: "The more things change, the more they stay the same."[74] Before there were AI-assisted deepfakes, and even now, "living deepfakes" sometimes make deceptive appearances among us. For example, there is the legendary 2004 Saint Augustine, Florida appearance of a Ringo Starr look-alike who toured a local museum, played the piano in the ballroom of the Casa Monica Hotel, and simply uttered "no comment" when asked what he was doing in Saint Augustine.[75] On August 16, 2025, a convincing Justin Bieber impersonator hoaxed his way into performing an impromptu set at the XS Nightclub at the Wynn Hotel in Las Vegas.[76] The author of this article was less than thrilled (but begrudgingly amused) when the "French Elvis" sang a pretty convincing version of *Love Me Tender* to the author's wife outside the gate to Graceland in Memphis, Tennessee – delaying the author's more highly anticipated visit to nearby Nesbit, Mississippi to meet the real Jerry Lee Lewis.[77] Years ago, as described in *Allen v. National Video, Inc*., a photograph of a Woody Allen impersonator with an uncanny talent for

---

be found in, *e.g.,* Wiener, *The Human Use of Human Beings, Cybernetics And Society*, Da Capo Press, 1988 (reprint; originally published in 1950 by The Riverside Press, Cambridge, Massachusetts; revised in 1954). See especially the section of the book, *Some Communication Machines and Their Future.*

[72] Seriously, does anybody really like these long, acronym-driven titles?

[73] Hammurabi's Code (circa 1780 BCE) has been described as follows: "One of the most influential codifications of law in ancient history, the text provides students with a concrete example of the expanding influence of centralized government on the personal and professional lives of the general population." Stockdale, *World History Sources*, Course Description, Department of History, University of Central Florida, https://chnm.gmu.edu/worldhistorysources/d/267/whm.html. The first law in the Code is: "If any one ensnare another, putting a ban upon him, but he cannot prove it, then he that ensnared him shall be put to death." The Code of Hammurabi, Translated by L.W. King, Yale Law School, The Avalon Project, https://avalon.law.yale.edu/ancient/hamframe.asp.

[74] According to one source: "The first recorded use of this expression is by French critic, journalist and novelist Alphonse Karr in 1849 in *Les Guêpes*, a monthly journal he founded: *Plus ça change, plus c'est la même chose*." BookBrowse, https://www.bookbrowse.com/expressions/detail/index.cfm/expression_number/483/the-more-things-change-the-more-they-stay-the-same; Some sources also point to Ecclesiastes 1:2-11 as expressing (roughly) the same sentiment. *See,* Enter the Bible, *Ecclesiastes 1:2-11 – The more things change, the more they stay the same,* https://enterthebible.org/passage/ecclesiastes-12-11-the-more-things-change-the-more-they-stay-the-same#:~:text=Ecclesiastes%201%3A2%2D11%20%E2%80%93,more%20they%20stay%20the%20same.

[75] *See* Romenesko, *Oops!: St. Augustine Record is fooled by Ringo look-alike*, Poynter.50, Aug. 11, 2004, https://www.poynter.org/reporting-editing/2004/oops-st-augustine-record-is-fooled-by-ringo-look-alike/.

[76] Blair, *Justin Bieber impersonator tricks Vegas club into letting him perform, runs up $10K bar tab*, New York Post, Aug. 19, 2025, https://nypost.com/2025/08/19/us-news/justin-bieber-impersonator-tricks-vegas-nightclub-into-letting-him-perform/.

[77] Sorry for oversharing. I'm over it. Really.

mimicry appeared without Allen's consent in an advertisement for a nationally franchised video rental chain.[78] More recently, as described in *Forrest v. Meta Platforms, Inc.,* Dr. Andrew Forrest, a prominent Australian businessman and philanthropist brought a case against Meta Platforms, Inc. based on its publication of alleged deepfake videos of Dr. Forrest endorsing cryptocurrency scams and other fraudulent investment products.[79]

Sometimes, "real" imposters and AI-assisted deepfakes raise all sorts of economic issues, general privacy issues, intellectual property issues, personal and business tort issues, criminal law issues, deceptive practices issues, and contract considerations, such as the Guild agreement provisions that govern the use of AI deepfakes in motion pictures and video games.[80] These types of scenarios and issues (some of which involve urgent public safety considerations) merit immediate attention, but for now the focus of the AI Action Plan is on "nonconsensual intimate visual depictions"[81] and the pollution of our judicial and other evidence-driven processes by synthetic media. As the discussion above concedes, the presence of deceit and fakery among us is nothing new. But it is the generally accessible AI technology and Internet technologies to readily create and rapidly distribute the fake content that heightens the concern and the amount of injury that can be inflicted – and the cases are starting to multiply.[82]

II. A Nudge To Congress: Lend Me Your Ears - The Problem Of Audio Digital Forgeries

[78] *Allen v. National Video, Inc*., 610 F. Supp. 612 (S.D.N.Y. 1985). For those of us born before the Internet, we immediately recall that "brick and mortar" video rental stores provided hard copies of movies – in the form of VHS tapes and then DVDs – for rental (constantly reminding us to be courteous and rewind rented VHS tapes). For those born too late to experience browsing and actually holding the art in a massively-stocked retail setting, the closest experience today is provided by some local libraries (like the one in Ormond Beach, Florida) that still stock DVDs for patrons like me. OK, I confess to a preference for hard copy books and vinyl records too. But, on a more practical level, when was the last time somebody hacked your computer through your vinyl record player?

[79] *Forrest v. Meta Platforms, Inc.,* 737 F.Supp.3d 808 (N.D. Cal. 2024).

[80] *See, e.g.,* SAG-AFTRA, *Regulating Artificial Intelligence*, TV/THEATRICAL, 2023, https://www.sagaftra.org/sites/default/files/sa_documents/AI%20TVTH.pdf; SAG-AFTRA, 2025 Interactive Media Video Game Agreement, https://www.sagaftra.org/contracts-industry-resources/interactive/2025-interactive-media-video-game-agreement.

[81] The Act incorporates the definition of "intimate visual depiction" that appears in section 1309 of the Consolidated Appropriations Act, 2022 (15 U.S.C. §6851), the definition therein being: "The term "intimate visual depiction"— (A) means a visual depiction, as that term is defined in section 2256(5) of title 18 that depicts— (i) the uncovered genitals, pubic area, anus, or post-pubescent female nipple of an identifiable individual; or (ii) the display or transfer of bodily sexual fluids— (I) on to any part of the body of an identifiable individual; (II) from the body of an identifiable individual; or (III) an identifiable individual engaging in sexually explicit conduct and (B) includes any visual depictions described in subparagraph (A) produced while the identifiable individual was in a public place only if the individual did not— (i) voluntarily display the content depicted; or (ii) consent to the sexual conduct depicted." The cited definition of "visual depiction: is: "visual depiction" includes undeveloped film and videotape, data stored on computer disk or by electronic means which is capable of conversion into a visual image, and data which is capable of conversion into a visual image that has been transmitted by any means, whether or not stored in a permanent format."

[82] *See, e.g., Spone v. Reiss,* 2025 WL 670968 (3rd Cir. 2025); *Broadrick v. Gilroy*, No. 3:24-cv-1772 (VAB), ___ F.Supp.3d ___(C.D. Conn. 2025) ; *Doe ONE v. The Nature Conservancy*, 2025 WL 1232527 (D. Minn. 2025).

As noted in the AI Action Plan, the recently enacted TAKE IT DOWN Act helps to protect against the nonconsensual publication of intimate images, including authentic materials and digital forgeries.[83] While the Act's definition of affected platforms includes "a website, online service, online application, or mobile application that serves the public and primarily provides a forum for user-generated content, including messages, videos, images, games, and ***audio*** files"[84] (emphasis supplied), the Act focuses on nonconsensual publication intimate visual depictions – or in the case of minors, situations in which consent cannot be given. Therefore, it seems that AI-assisted or simulated audio recording of an intimate nature, via "voice cloning," or nonconsensual publication of intimate audio episodes, will (for now) likely be treated under other laws – such as state legislation, other federal statutes, or tort and criminal law principles – or under an amended or judicially clarified TAKE IT DOWN Act if the problem manifests itself with growing force.

As demonstrated in *Lehrman v. Lovo, Inc*., the courts are beginning to see "voice cloning" claims in a number of contexts – for example, copyright, Lanham Act false association claims, rights of publicity (*e.g.,* New York Civil Rights Law Section 50), contract, consumer protection law, and fraud.[85] *Lehman* involved an allegation that two actors' voices were used without the actors' permission to train a generative AI system – the result being advertisements and other audio content that simulated the actors' voices but were not actually supplied by or consented to by the actors. It is not difficult to imagine what other types of nonconsensual uses of identifiable deepfake audio are already out there or waiting in the wings. Also, in a somewhat analogous evolution, early rights of publicity cases tended to focus on photographs and images,[86] but the concept of proprietary rights in identifiable voices[87] used without permission developed naturally

---

[83] AI Action Plan, p 12. The TAKE IT DOWN Act, also titled *Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act*, prohibits the nonconsensual online publication of intimate visual depictions of individuals, both authentic and computer-generated, and requires certain online platforms to promptly remove such depictions upon receiving notice of their existence. *See* https://www.govtrack.us/congress/bills/119/s146. The Act was introduced by Senator Red Cruz (R-Texas) and Senator Amy Klobuchar (D-Minnesota), championed by First Lady Melania Trump, and was signed into law by President Trump on May 19, 2025. *See* https://www.congress.gov/bill/119th-congress/senate-bill/146/text. A recent discussion of the TAKE IT DOWN Act can be found at Sundberg*, Federalizing NCII Regulations: The Take It Down Act's Approach To Criminalization, Platform Liability, And Threats To Disseminate,* Georgia Law Review, Vol. 59, Issue 3, 2025, https://georgialawreview.org/wp-content/uploads/2025/07/Trudi-Sundberg_Federalizing-NCII-Regulation.pdf.

[84] *See* Act, Sec. 4, DEFINITIONS, (3)(A)-(B), https://www.congress.gov/bill/119th-congress/senate-bill/146/text.

[85] *See Lehrman v. Lovo, Inc.*, 2025 WL 1902547 (S.D.N.Y. 2025).

[86] *See, e.g., Pavesich v. New England Life Insurance Co.,* 122 Ga. 190, 50 S.E. 68 (1905)(unauthorized use of plaintiff's photograph to advertise insurance); *Edison v. Edison Polyform Mfg. Co.,* 67 A. 392 (N.J. Ch. 1907)(unauthorized use of Thomas Edison's image on product label); *see also Roberson v. Rochester Folding Box Co.* 171 N.Y. 538, 64 N.E. 442 (N.Y. 1902)(unauthorized use of plaintiff's image on bags of flour and advertisement for the flour). The denial of relief in the *Roberson* case led to New York's enactment of a civil rights law to product against the unauthorized use of an individual's photographic likeness to advertise products.

[87] *See, e.g., Waits v. Frito-Lay, Inc.*, 978 F.2d 1093 (9th Cir. 1992)(unauthorized imitation of plaintiff's distinctive voice in advertisement for snack food); *Midler v. Ford Motor Co., 849 F.2d 460* (9th Cir. 1988) (unauthorized imitation of plaintiff's distinctive voice in advertisement for automobiles). Notably, each of the cases used reenforcing content – such as a song associated with the plaintiff – to bolster the recognizability of the imitation. Similarly, audio deepfakes can use such reinforcing cues as a person's name, a celebrity's catchphrase, famous lines

from this core concern regarding rights in identity. It seems that AI-assisted audio fakes may be the next potential "frontier" for TAKE IT DOWN type treatment.[88]Moreover, it should come as no surprise that Hollywood's largest actor's union, SAG-AFTRA, has already focused on voice cloning and has prepared agreement terms, conditions, and controls for its members.[89]

III. The Malevolence Of Indifference – Large Language Models And Injurious Fakery

In addition to audio constructs, potentially harmful Large Language Models ("LLMs")[90] – including LLM-powered chatbots[91] – have been the subjects of urgent consideration.[92] Alleged LLM injury can range from a false representation of the chatbot (including "seductive" bots) as a specific, actual individual (whose consent might not have been secured) to allegations that a chatbot contributed to the user's act of murder and/or suicide[93] – and then there is the recent situation where, according to (possibly histrionic) media reports, "[a] 76-year-old retiree

---

spoken in a movie in which the celebrity appeared, or lyrics from a song made famous by the celebrity. Of course, this list is not exhaustive.

[88] The author notes that the topic of unauthorized audio and the potential application of federal TAKE IT DOWN types of criminal provisions and removal mechanisms merit in-depth research and discussions that are beyond the scope of this article. So too, the topic of malicious and injurious chatbots is urgent and needs additional, serious attention.

[89] *See* Whitfill Roeloffs, *Some Actors Will Let AI Replicate Their Voices For Advertisements Under New Union Agreement*, Forbes, Aug. 14, 2024, https://www.forbes.com/sites/maryroeloffs/2024/08/14/some-actors-will-let-ai-replicate-their-voices-for-advertisements-under-new-union-agreement/.

[90] "Large language models (LLMs) are a category of foundation models trained on immense amounts of data making them capable of understanding and generating (what appears to be) natural language and other types of content to perform a wide range of tasks . . . In a nutshell, LLMs are designed to understand and generate text like a human, in addition to other forms of content, based on the vast amount of data used to train them. They have the ability to infer from context, generate (at times) coherent and contextually relevant responses, translate to languages other than English, summarize text, answer questions . . . and even assist in creative writing or code generation tasks." *See, What are LLMs*, IBM, https://www.ibm.com/think/topics/large-language-models, (parenthetical matter added).

[91] For a concise discussion of LLMs and chatbots, *see* Kuka, *Differences Between Chatbots and LLMs*, Learn Prompting, March 6, 2025, https://learnprompting.org/docs/basics/chatbot_basics?srsltid=AfmBOoqmz1Hkdf8vExViK61kT0K6-HvcPt376zefZwW6tiJ7sGHnAH9Z.

[92] *See, Senator Padilla Introduces Legislation To Protect Children From Predatory Chatbot Practices*, Press Release, Feb. 3, 2025, https://sd18.senate.ca.gov/news/senator-padilla-introduces-legislation-protect-children-predatory-chatbot-practices. As noted by Senator Padilla, a state senator representing California's District 18, "Our children are not lab rats for tech companies to experiment on at the cost of their mental health. We need common sense protections for chatbot users to prevent developers from employing strategies that they know to be addictive and predatory." *Id. See also,* Rieper, *State Lawmakers Propose Regulating Chatbots*, Multistate, Jan. 24, 2025, https://www.multistate.ai/updates/vol-46; *Critical Assembly Committee Advances Legislation Protecting Against Predatory Chatbot Practices*; California State Senator Steve Padilla Press Release, July 8, 2025, https://sd18.senate.ca.gov/news/critical-assembly-committee-advances-legislation-protecting-against-predatory-chatbot.

[93] *See* Zilber, *How ChatGPT fueled delusional man who killed mom, himself in posh Conn. Town*, New York Post, Aug. 29, 2025, https://nypost.com/2025/08/29/business/ex-yahoo-exec-killed-his-mom-after-chatgpt-fed-his-paranoia-report/; Payne*, An AI chatbot pushed a teen to kill himself, a lawsuit against its creator alleged,* AP, Oct. 25, 2024, https://apnews.com/article/chatbot-ai-lawsuit-suicide-teen-artificial-intelligence-9d48adc572100822fdbc3c90d1456bd0.

from New Jersey met a tragic end from a fall while trying to meet up with a flirty Kendall Jenner[94] lookalike called 'Big sis Billie' – without realizing "she" was an AI chatbot."[95] As one set of commentators on the dangers of seductive AI behavior – and potentially addictive sycophancy in AI systems – noted:

> AI companionship is no longer theoretical—our analysis of a million ChatGPT interaction logs reveals that the second most popular use of AI is sexual role-playing. We are already starting to invite AIs into our lives as friends, lovers, mentors, therapists, and teachers. . . . Addressing the harm that AI companions could pose requires a thorough understanding of the economic and psychological incentives pushing forward their development. Until we appreciate these drivers of AI addiction, it will remain impossible for us to create effective policies.[96]

In the history of scientific and technological development, we have (in a real and metaphoric sense) invented means to start fires and then couldn't resist the urge to play with them – now, the technology is not only playing back, it has developed an independent expertise in doing so.[97] This creates the need for recognition, regulation, positive design approaches, and public education about the dangers of "shared hallucinations" with AI, especially LLMs.

A recent decision that sent shockwaves through developers and suppliers of chatbot-app technologies is *Garcia v. Character Technologies, Inc*. – a case that involved an app called Character A.I.[98] In *Garcia*, a mother, individually and as the personal representative of the estate of her teenage son who allegedly died by bot-influenced suicide, withstood parts of a motion to dismiss as the Court held that the plaintiff:

> (1) adequately alleged the defendant company is an alter ego of its nonresident founders, as basis for personal jurisdiction;
>
> (2) adequately alleged that the supply of component parts provided a basis for product liability claims against the provider;

---

[94] For those who might not be "up to speed" on celebrity Kendall Jenner, she has been described as "an American socialite, television personality and model." This summary, and more biographical information, can be found at Kendall Jenner Biography, IMBb, https://www.imdb.com/name/nm2832525/bio/.

[95] *Id. See also,* Taylor, *She looked like Kendall Jenner, whispered sweet nothings… and lured a New Jersey retiree to his death,* Daily Mail, Aug. 16, 2025, https://www.dailymail.co.uk/news/article-15006229/kendall-jenner-ai-chatbot-lured-new-jersey-retiree-death.html; Davies, *Is this AI video ACTUALLY Kendall Jenner? Thousands duped by Meta bot that experts insist is the REAL celebrity*, Daily Mail, Oct. 15, 2023, https://www.dailymail.co.uk/sciencetech/article-12622719/Kendall-Jenners-alter-ego-duping-fans-AI-generated-posts.html.

[96] Mahari & Pataranutaporn, *We need to prepare for 'addictive intelligence,'* MIT Technology Review, Aug. 5, 2024, https://www.technologyreview.com/2024/08/05/1095600/we-need-to-prepare-for-addictive-intelligence/.

[97] A discussion of LLM sycophancy can be found in Sponheim, *Sycophancy in Generative-AI Chatbots*, NN/g, Jan. 12, 2024, https://www.nngroup.com/articles/sycophancy-generative-ai-chatbots/.

[98] *Garcia v. Character Technologies, Inc*., 2025 WL 1461721 (M.D. Fla. 2025).

(3) adequately pleaded claim against an investor for aiding and abetting the underlying product-liability tort;

(4) adequately alleged that the subject app was a defective product;

(5) adequately alleged that the defendants owed a duty to lessen the app's risk to users or see that sufficient precautions were taken;

(6) adequately pleaded a fraud-based claim under the Florida Deceptive and Unfair Trade Practices Act ("FDUTPA"); and

(7) adequately alleged that the investor obtained direct benefit from teenager, as basis for unjust enrichment claim.[99]

Notably, when confronted with an argument that Character A.I.'s output is speech protected by the First Amendment, the *Garcia* Court responded that, at this early stage in the case, issues regarding whether the output actually constitutes speech protected by the First Amendment precluded dismissal.[100] Regarding whether the subject app is actually a defective product (rather than a service) the Court noted that ideas, images, information, words, expressions, and concepts are generally not recognized as products for purposes of applying product liability law.[101] However, the Court also noted that there is nonuniformity in courts' determinations on whether virtual platforms, such as social media sites, are products.[102] Ultimately, the Court stated that: "Character A.I. is a product for the purposes of Plaintiff's product liability claims so far as Plaintiff's claims arise from defects in the Character A.I. app rather than ideas or expressions within the app."[103] These preliminary determinations, plus, *e.g.,* the Court's willingness to entertain claims regarding corporate alter ego liability, investor liability, "component" supplier liability, and jurisdiction, appear to have established some "battle lines" in cases involving

[99] *Id.* at pp. 1-2.
[100] *Id.* at p. 13.
[101] *Id. citing Wilson*, 198 F. Supp. 2d at 170, 173 (finding that a video game, which the plaintiff alleges inspired a player to stab her son, was not a product because the harm resulted from the intangible expressive ideas of the video game); *Watters v. TSR, Inc.*, 904 F.2d 378, 381 (6th Cir. 1990) (declining to extend strict liability "to words or pictures" in Dungeons and Dragons literature) and noting the statement that courts "separate the sense in which the tangible containers of [ ] ideas are products from their communicative element for purposes of strict liability." *James v. Meow Media, Inc.*, 300 F.3d 683, 701 (6th Cir. 2002) (finding that "the ideas conveyed by the video games, movie cassettes and internet transmissions," which the plaintiff alleges "caused [a consumer] to kill his victims," was not a product).
[102] *Id.* noting that *Jacobs v. Meta Platforms, Inc.*, No. 22-cv-5233, 2023 WL 2655586, at *4 (Cal. Super. Mar. 10, 2023) (finding that "as a social media platform that connects its users, Facebook is more akin to a service than a product," but not considering whether the platform's "recommendation algorithms or related features, such as newsfeeds or those related to social groups, may be considered 'products' "), can be compared with *In re Soc. Media Adolescent Addiction/Pers. Inj. Prods. Liab. Litig.*, 702 F. Supp. 3d 809, 849, 854 (N.D. Cal. 2023) (finding that the alleged defects in the functionalities of the defendants' social media platforms were "analogizable to tangible personal property" rather than "akin to ideas, content, and free expression" and could thus support a claim for product liability).
[103] *Id.* at p. 14.

developers and suppliers of bot technologies and their potential defenses against users and others who allegedly suffer injury from those technologies.[104]

There are continuing attempts and evolving strategies to ensure that LLMs – including LLM-driven chatbots – are checked and sanitized to avoid injurious output. Among such problem identification or "sanitizing" techniques are machine unlearning[105] and the use of constructive adversarial jailbreaking[106] to help test and ensure appropriate bot or resource behavior – but the efficacy of such approaches have been questioned by skeptics[107] Nonetheless, a recent alert in a discussion of potentially harmful LLMs is as follows: ***If you or someone you know may be experiencing a mental-health crisis or contemplating suicide, call or text 988. In emergencies, call 911, or seek care from a local hospital or mental health provider. For listing of local or international resources please see https://www.iasp.info/suicidalthoughts/.***[108]

In order to revisit and achieve some "closure" on the Kendall Jenner matter it should be noted that there are a number of other celebrities who also licensed their images for use as fictional character chatbot interfaces – including Tom Brady, Paris Hilton, and Snoop Dog[109] – and these "cultural icons" appeared as the fictionalized character faces of responsive AI systems on

---

[104] For another interesting decision on chatbot communications as a basis for jurisdiction, *see Barton v. Pinnacle Home Improvements, LLC,* 2024 WL 2943805 (W.D. Wash. 2024)(defendant Georgia company that used a chatbot knew or should have known the substance of Washington state resident's communication seeking discontinuation of solicitations, and defendant's telemarketing messages to him after that point were purposefully directed at and engendered jurisdiction in Washington state.).

[105] Machine unlearning is a technique that forces LLMs to forget problematic training instances and thereby minimize their influence. *See* Song, Kim, Kim, Shin, and Son, *Refusal Is Not an Option: Unlearning Safety Alignment of Large Language Models*, Paper presented at USENIX (The Advanced Computing Systems Association) Aug. 13-15, 2025. https://www.usenix.org/system/files/usenixsecurity25-song-minkyoo.pdf. The cited paper comprises an interesting discussion of "novel attack methods designed to break LLM safety alignment through unlearning."

[106] In general, "jailbreaking" refers to the use of "prompts" or inputs designed to cause models to deviate from their safety filters and operational guidelines. The technique can be used to identify vulnerabilities in a system. For a discussion of jailbreaking, *see, e.g.,* Awoufack, *Adversarial Prompt Transformation for Systematic Jailbreaks of LLMs*, Thesis submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of Master Of Engineering In Electrical Engineering And Computer Science at the Massachusetts Institute Of Technology, 2024. https://dspace.mit.edu/bitstream/handle/1721.1/157167/awoufack-awoufack-meng-eecs-2024-thesis.pdf?sequence=1&isAllowed=y. For an additional discussion of various techniques that reveal vulnerabilities and dangers in LLM prompting, *see Adversarial Prompting in LLMs*, (updated July 5, 2025), https://www.promptingguide.ai/risks/adversarial.

[107] *See* Schoene & Canca, '*For Argument's Sake, Show Me How To Harm Myself! Jailbreaking LLMs In Suicide And Self-Harm Contexts,* PREPRINT, Institute for Experimental AI, Northeastern University, July 8, 2025, https://arxiv.org/pdf/2507.02990. One of the results claimed in the Schoene & Canca study is that: "[W]e evaluated six widely available LLMs on two test cases in mental health—namely self-harm and suicide. We show that despite existing safety features and guardrails, LLMs still output potentially harmful content despite being prompted for the user's previously disclosed intent to cause harm." *Id.* at p. 7. **ALSO, SEE THE MENTAL HEALTH ASSISTANCE NOTICE IN THE MAIN BODY OF THIS ARTICLE.**

[108] *See, e.g.,* de Guzman, *AI Chatbots Can Be Manipulated to Provide Advice on How to Self-Harm, New Study Shows,* Time, July 31, 2025, https://time.com/7306661/ai-suicide-self-harm-northeastern-study-chatgpt-perplexity-safeguards-jailbreaking/.

[109] Davis, *Kendall Jenner's A.I. Clone Has Been Terminated, and I Am Sad*, artnet, July 31, 2024, https://news.artnet.com/art-world/kendall-jenner-a-i-clone-terminated-meta-2518771.

WhatsApp, Messenger, and Instagram, along with character backstories on Facebook.[110] This AI-powered celebrity chatbot project was eventually terminated by its provider, Meta, due to a lack of public receptivity to Meta's approach[111] – but the practical considerations and the terms of celebrity license agreements for LLM applications that might generate unanticipated content still present interesting issues and challenges in, *e.g.,* the areas of agreement terms and technological controls.[112] Of course, getting back to the TAKE IT DOWN Act's limitations, an image of an individual that is not an within the Act's definition of "intimate visual depiction" can be appropriated by a chatbot's provider to add an attractive face to an app/bot that provides injurious, salacious or otherwise offensive content.

The potential "AI Universe" candidates for specific and detailed federal and state regulation seem to be lining up as the present lack of coordinated or effective controls – and those candidates' injurious consequences – become more widely recognized. However, as noted in a recent Congressional Research Service discussion:

> The approach of the U.S. federal government as a whole appears to be cautious in regard to regulating AI in the private sector and more focused on oversight of federal government uses of AI. In the absence of federal AI regulations, states have been enacting their own laws. Critics assert that such a patchwork of AI laws creates challenges for companies and that a nationwide regulatory structure may incentivize product development.[113]

An additional consideration is that ill-conceived, under-researched, or overreaching federal or state regulation can damage an emergent industry which will (at least in some forms) be vital to preserving and advancing U.S. interests.

So, what are some important details regarding the new "nationwide" federal legislation that addresses nonconsensual intimate visual depictions?

---

[110] Meta, *Introducing New AI Experiences Across Our Family of Apps and Devices*, September 27, 2023, https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/.

[111] *See* Zilber, *Meta scraps failed celebrity AI chatbots after users ignored them, deemed them creepy*, New York Post, Aug. 1, 2024, https://nypost.com/2024/08/01/business/meta-ends-failed-celebrity-ai-chatbots-after-users-ignored-them/.

[112] In one of the most notorious "old school" examples of an actor being blindsided by unanticipated content associated with his name and performance occurred when actor Malcolm McDowell discovered that the version of the film *Caligula* that he signed up for and acted in was then turned into what many viewed as a pornographic travesty by, *e.g.,* the producer's later insistence on adding graphic, gratuitous, and arbitrary sex scenes. *See, e.g.,* McAndrews*, Malcolm McDowell Shares His Experience Watching 'Caligula – The Ultimate Cut' On The Big Screen [Fantastic Fest 2023]*, Dread Central, Oct. 9, 2023, https://www.dreadcentral.com/interviews/464788/malcolm-mcdowell-shares-his-experience-watching-caligula-the-ultimate-cut-on-the-big-screen-fantastic-fest-2023/. This bit of history from Hollywood serves not only as the story of a bad movie, but as a cautionary tale for actors and other celebrities contemplating an unpredictable LLM-driven celebrity dialog or animation project.

[113] Harris, *Summary - Regulating Artificial Intelligence: U.S. and International Approaches and Considerations for Congress*, Congressional Research Service, R48555m June 4, 2025, https://www.congress.gov/crs_external_products/R/PDF/R48555/R48555.2.pdf.

IV. The TAKE IT DOWN ACT – Finally

A. Introduction

The TAKE IT DOWN Act, which was signed into law by President Trump on May 19, 2025, makes two key changes to federal law: (1) it amends Section 223 of the Communications Act of 1934[114] to add new criminal prohibitions related to the intentional disclosure of nonconsensual intimate visual depictions; and (2) it imposes new requirements for covered platforms to establish a specified notice-and-removal process by May 19, 2026.[115] Introduced by Senator Ted Cruz (R-Texas), the Act had twenty-one cosponsors, including Senator Amy Klobuchar (D-Minnesota),[116] passed in the Senate by unanimous vote,[117] and passed in the House of Representatives with a vote of 409, with two "no" votes[118] — and twenty-two not voting.[119] Congresswoman Maria Elvira Salizar,[120] a staunch advocate for the Act, noted that over 120 organizations representing victim advocacy groups, law enforcement, and leaders in the tech industry voiced their support for the TAKE IT DOWN Act, including Meta, Snap, Google, Microsoft, TikTok, X, Amazon, Bumble, Match Group, Entertainment Software Association, IBM, TechNet, the U.S. Chamber of Commerce, Internet Works, the National Fraternal Order of Police, the National Center for Missing and Exploited Children (NCMEC), RAINN (Rape, Abuse & Incest National Network), and the National Center on Sexual Exploitation (NCOSE).[121]

---

[114] *See* 47 U.S.C. § 223.

[115] Killion, *The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images*, Congress.Gov, Congressional Research Service, https://www.congress.gov/crs-product/LSB11314.

[116] *See* Congress.Gov, S.146 – TAKE IT DOWN Act, https://www.congress.gov/bill/119th-congress/senate-bill/146/cosponsors.

[117] *See* Congress.Gov. S. 146 – TAKE IT DOWN Act, https://www.congress.gov/bill/119th-congress/senate-bill/146/all-actions?overview=closed#tabs,

[118] Eric Burlison (R-Missouri) and Thomas Massie (R-Kentucky) were the "no" votes, *See* Congress.Gov, House Roll Call Vote, https://www.congress.gov/votes/house/119-1/104. A spokesperson for Rep. Burlison explained the Congressman's "no" vote, in part, as follows: "Sharing non-consensual intimate imagery is abhorrent, but this bill unnecessarily federalizes the criminalization of conduct that states already have laws against, and many of them already have laws tailored to address AI-generated deepfakes. As such, creating a new federal offense is both redundant and constitutionally problematic. The further federalization of criminal law in this way undermines state authority, blurs lines of accountability, and risks duplicative prosecutions. Furthermore, while the bill attempts to safeguard free speech with a 'reasonable person' test, I remain concerned about its impact on First Amendment rights and the unchecked growth of federal power." *see Padgett, "Take it Down Act" heads to Trump's Desk; Rep. Burlison explains why he voted no,* Ozarks First, Updated May 5, 2025, https://www.ozarksfirst.com/news/trump-signs-revenge-porn-bill/; Congressman Massie's April 29, 2025 Facebook post on his "no" vote states: "Last night we voted on the "TAKE IT DOWN Act," a bill that would impose federal criminal and civil penalties for publishing unauthorized intimate pictures generated with AI. I voted NO because I feel this is a slippery slope, ripe for abuse, with unintended consequences." https://www.facebook.com/RepThomasMassie/posts/last-night-we-voted-on-the-take-it-down-act-a-bill-that-would-impose-federal-cri/1227366282079363/.

[119] *See* Congress.Gov. S. 146 – TAKE IT DOWN Act, https://www.congress.gov/bill/119th-congress/senate-bill/146/all-actions?overview=closed#tabs,

[120] Representative from Florida's 27th District. *See* https://salazar.house.gov/.

[121] *See* Salazar, *Take It Down Act Passes the House and Heads to President's Desk*, Press Release, April 28, 2025, https://salazar.house.gov/media/press-releases/take-it-down-act-passes-house-and-heads-presidents-desk. A good

Paris Hilton – a victim of unauthorized publication of intimate images – voiced her strong support for the Act in a posting on X.[122] Also, First Lady Melania Trump was exceptionally active in supporting the Act from its inception to its final enactment.[123] In short, the Act received exceptionally strong bi-partisan, celebrity, and public support. Notably, this vigorous and coordinated campaign to address the crucial issue of nonconsensual intimate visual depictions, and the particular dangers of AI in this context, can serve as a model (or inspiration) for further bi-partisan drives to confront other species of injurious AI technologies and uses.

Notably, the Act does not expressly include a separate private right to sue – and at least one commentator notes that there are potential alternative arguments that: (1) the private right of action under the 2022 reauthorization of the Violence Against Women Act (VAWA) [124] for unaltered nonconsensual intimate imagery (NCII) might be expanded to include "digital forgeries" when interpreted in light of the TAKE IT DOWN Act's verbiage; or (2) "because the Take It Down Act creates different offenses for unadulterated NCII versus digital forgeries, and because Congress did not amend the VAWA to specifically include digital forgeries, courts may infer that Congress intended the private right of action to apply only to unadulterated NCII." [125] In any case, VAWA's application to digital forgeries is currently unsettled.[126] Therefore, it appears that the limitations in the Act comprise a lost opportunity to clarify and strengthen the ability of individuals to advance claims based on digital forgeries. Still, in light of the advances provided by the Act – and the indication of a broad legislative awareness of the dangers of AI – it is likely

---

compilation of statements of support for the pending legislation can be found at *What They're Saying/The TAKE IT DOWN Act*, https://www.commerce.senate.gov/services/files/3EED5B40-FE58-4799-A191-F477C147A7C0.

[122] The text of Hilton's message of support is as follows: "Today the House Energy and Commerce Committee will be voting on the Take It Down Act. This bill has already passed the Senate! No more excuses, loopholes, or delays—victims deserve justice, privacy, and protections. I'm proud to stand with this brave community - together, we can turn pain into purpose and protect future generations from this harm. #TAKEITDOWNAct #LegislationisHot."1:17 PM – March 26, 2025, https://x.com/ParisHilton/status/1904945762385142228.

[123] *See, e.g., First Lady Melania Trump Joins President Trump for Signing of the "Take It Down" act*, The White House, May 19, 2025 in which the following (excerpted) quote appears: "This legislation is a powerful step forward in our efforts to ensure that every American—especially our young people—can feel better protected from their image or identity being abused through non-consensual intimate imagery or NCII. Artificial Intelligence and social media are the digital candy of the next generation—sweet, addictive, and engineered to have an impact on the cognitive development of our children. But unlike sugar, these new technologies can be weaponized, shape beliefs, and sadly, affect emotions and even be deadly." https://www.whitehouse.gov/briefings-statements/2025/05/first-lady-melania-trump-joins-president-trump-for-signing-of-the-take-it-down-act/.

[124] *Citing* 15 U.S.C. § 6851. On March 15, 2022, Congress authorized a federal civil claim relating to the disclosure of intimate images as part of the Consolidated Appropriations Act, 2022. The new cause of action, which takes effect on October 1, 2022, marks the first federal law targeting the unauthorized dissemination of private, intimate images of both adults and children—images commonly referred to as "nonconsensual pornography" or "revenge porn." *See,* Killion, *Federal Civil Action for Disclosure of Intimate Images: Free Speech Considerations*, Congress.Gov, April 1, 2022, https://www.congress.gov/crs-product/LSB10723.

[125] *See* Leibert, *Congress's Attempt to Criminalize Nonconsensual Intimate Imagery: The Benefits and Potential Shortcomings of the TAKE IT DOWN Act*, National Association of Attorneys General, Aug. 26, 2025, https://www.naag.org/attorney-general-journal/congresss-attempt-to-criminalize-nonconsensual-intimate-imagery-the-benefits-and-potential-shortcomings-of-the-take-it-down-act/.

[126] *Id., citing* Victoria L. Killion, The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images (May 20, 2025), https://www.congress.gov/crs-product/LSB11314.

best to take a "glass half full" view of the situation and advocate for further affirmative legislative developments and clarifications. Also, as discussed below, the criminal penalties included in the Act include restitution as under section 2264 of title 18, United States Code. This potentially gives some comfort to victims, but the ability to personally advance a claim under a clear, well-constructed, coherent, and uniform civil law with appropriate jurisdiction – and with the possibility of enhanced damages – would provide an additional, desirable detriment and remedy to nonconsensual publication of intimate digital forgeries (or threats to do so).

B. The Act's Notice And Removal Provisions

In addition to its provision of criminal sanctions against violations, the Act recognizes individual rights and initiatives (in an arguably limited fashion) by providing a notice-and-removal process with regard to nonconsensual intimate visual depictions – and, as previously noted, covered platforms have until May 19, 2026 to implement the required notice-and-removal process.[127] Regarding platforms that will need to comply with the Act's notice and removal requirements, the Act's definition of "covered platform," as well as specific exclusions from the definition, are as follows:

> (A) In general. --The term ``covered platform'' means a website, online service, online application, or mobile application--
>
> (i) that serves the public; and
>
> (ii)(I) that primarily provides a forum for user-generated content, including messages, videos, images, games, and audio files; or
>
> (II) for which it is in the regular course of trade or business of the website, online service, online application, or mobile application to publish, curate, host, or make available content of nonconsensual intimate visual depictions.
>
> (B) Exclusions. --The term ``covered platform'' shall not include the following:
>
> (i) A provider of broadband internet access service (as described in section 8.1(b) of title 47, Code of Federal Regulations, or successor regulation).
>
> (ii) Electronic mail.
>
> (iii) Except as provided in subparagraph (A)(ii)(II), an online service, application, or

[127] *See* Act, Sec. 3, NOTICE AND REMOVAL OF NONCONSENSUAL INTIMATE VISUAL DEPICTIONS, (a)(1)(A), https://www.congress.gov/bill/119th-congress/senate-bill/146/text.

website—

(I) that consists primarily of content that is not user generated but is preselected by the provider of such online service, application, or website; and

(II) for which any chat, comment, or interactive functionality is incidental to, directly related to, or dependent on the provision of the content described in subclause (I).[128]

While certain verbiage in the definition section – such as "*primarily* provides" in (ii)(I) (emphasis supplied) – may be subject to scrutiny, platforms may decide to " err on the side of inclusion" and look to the Act for guidance, even if the applicability of the Act is uncertain. However, in light of the definition's language and the specific exclusions, any characterizations of the Act as having an overbroad application to all websites that accept user input – and a general effect across the Internet – would be ill-founded. In any case, the covered platforms are required to provide "in easy to read and in plain language" a clear and conspicuous notice of the notice and removal process and its requirements.[129]

The Act authorizes the Federal Trade Commission to enforce compliance with the Act's notice-and-removal procedures with regard to both regular and nonprofit organizations[130] – noting that a failure to reasonably comply constitutes "a violation of a rule defining an unfair or deceptive act or practice" (UDAP violation) under the Federal Trade Commission Act.[131] Upon receipt of a proper notice, the recipient has forty eight (48) hours to: (a) remove the content; and (b) make reasonable efforts to identify and remove any known identical copies of such depiction.[132] In essence, a proper removal notice that satisfies the requirements of the Act[133] triggers a remedy akin to an *ex parte* injunction (with associated further investigative actions by the recipient). This forceful tool for automatic removal of content has drawn the ire of a number of organizations that view the Act's notice-and-removal process as violative of Free Speech, User Privacy, and Due Process – with one organization arguing that (especially via opportunities

[128] *See* https://www.congress.gov/bill/119th-congress/senate-bill/146/text.
[129] *See* Act, Sec. 3, NOTICE AND REMOVAL OF NONCONSENSUAL INTIMATE VISUAL DEPICTIONS, (a)(2)(A)-(B), https://www.congress.gov/bill/119th-congress/senate-bill/146/text.
[130] As noted by Killion, "[t]he Act extends the FTC's jurisdiction in this regard to nonprofit organizations, which are not usually covered by the FTC Act." *The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images*, Congress.Gov, Congressional Research Service, https://www.congress.gov/crs-product/LSB11314.
[131] *See* 15 U.S.C. § 57a(a)(1)(B).
[132] *See* Act, Sec. 3, NOTICE AND REMOVAL OF NONCONSENSUAL INTIMATE VISUAL DEPICTIONS, (a)(3)(A)-(B), https://www.congress.gov/bill/119th-congress/senate-bill/146/text.
[133] Section (a)(1)(i)-(iv) of the Act describes the required contents of a proper notice under the Act.

for frivolous or bad faith takedown requests) the process gives "the powerful a dangerous new route to manipulate platforms into removing lawful speech that they simply don't like."[134] Nonetheless, especially in light of the massive support received by the proposed legislation that matured into the Act, the prospect of abuse has been countered and outweighed by the damage nonconsensual intimate depictions can inflict on its victims. For example, in a March 3, 2025 Press Release from the U.S. Senate Committee on Commerce, Science, & Transportation, the rationale for the Act's notice-and-removal requirements was explained as follows:

> While nearly every state has a law protecting people from non-consensual intimate imagery (NCII), including 30 states with laws explicitly covering sexual deepfakes, these state laws vary in classification of crime and penalty and have uneven criminal prosecution. Further, victims struggle to have images depicting them removed from websites, increasing the likelihood the images are continuously spread and victims are traumatized.
>
> In 2022, Congress passed legislation creating a civil cause of action for victims to sue individuals responsible for publishing NCII. However, bringing a civil action can be incredibly impractical. It is time-consuming, expensive, and may force victims to relive trauma. Further exacerbating the problem, it is not always clear who is responsible for publishing the NCII.[135]

In short, an approach advocated by many supporters of the Act would be to address abuses of the Act if and as they arise, not to presume mass or extreme misuses and offer such presumptions as a basis to attack the Act and its protections against nonconsensual intimate depictions. Nonetheless, we can anticipate that the judicial system will have ample opportunities to consider whether the Act's notice-and-removal measures can withstand scrutiny under claims that free speech, user privacy, and whether due process are being imperiled or cast aside by the Act.

Anticipating that there might be claims against platforms for allegedly removing proper content under the Act's notice-and-removal procedures, the Act provides a "safe harbor" for covered platforms – sheltering them from liability "for any claim based on the covered platform's good faith disabling of access to, or removal of, material claimed to be a nonconsensual intimate visual depiction based on facts or circumstances from which the unlawful publishing of an intimate visual depiction is apparent, regardless of whether the intimate visual depiction is ultimately determined to be unlawful or not."[136] Notably, as stated, there is a "good faith"

---

[134] Kelley, *Congress Passes TAKE IT DOWN Act Despite Major Flaws*, Electronic Frontier Foundation, April 28, 2025, https://www.eff.org/deeplinks/2025/04/congress-passes-take-it-down-act-despite-major-flaws.

[135] *See* https://www.commerce.senate.gov/2025/3/house-leaders-pledge-to-advance-take-it-down-act-at-sen-cruz-s-bipartisan-roundtable-with-first-lady-melania-trump. The quoted section continues to list the following four key goals of the Act: (1) Criminalizing the publication of NCII in interstate commerce; (2) Protecting good faith efforts to assist victims; (3) Requiring websites to take down NCII upon notice from the victim; and (4) Protecting lawful speech. *Id.*

[136] *Id.*

requirement in the safe harbor provision. Therefore, as a practical matter, affected platforms will likely adopt standard operating procedures and best practices for exercising and documents the requisite good faith.

Also, an important point to note is that the TAKE IT DOWN Act's notice-and-removal requirements are somewhat reminiscent of the notice-and-takedown approach taken in the Digital Millennium Copyright Act (DMCA),[137] but the specific requirements under each law are not necessarily interchangeable.[138] Moreover platforms may be tempted to use or adapt the DMCA's copyright-oriented processes for notice and takedown procedures with regard to other intellectual property or tort claims, *e.g.,* trademark, defamation, privacy, etc.[139] This may be the case because the platform operators seek to establish that they have acted in good faith and reasonably with regard to such "non-copyright" matters – the argument being that the DMCA provided (at least inferentially) a standard for reasonable diligence. Also, it may be a matter of conservation of resources to consolidate such notice and takedown procedures under the platform's already-established DMCA mechanisms. Finally, many platforms use a policy and terms violation notice system to address non-copyright matters.[140] In other words, the platform's access and use terms govern users' behavior – and violations of the terms can be reported and used as a basis for, *e.g.,* removing content and blocking a violator's access to the site. However, as noted – and as a practical matter – separate (and serious) consideration should be focused on whether a pre-existing framework is suited to the specific requirements of the TAKE IT DOWN Act's notice and removal provisions.

C. Crime And Punishment

Unlike the Act's notice and removal requirements, the Act's criminal prohibitions took effect immediately when the Act was signed into law. As shown below, the Act's section on criminal activity distinguishes between violative depictions of adults and minors (*i.e.,* individuals under the age of 18 years) and also distinguishes between an unaltered image and a "digital forgery" – which means:

> "[A]ny intimate visual depiction of an identifiable individual created through the use of software, machine learning, artificial intelligence, or any other computer-generated or technological means, including by adapting, modifying, manipulating, or altering an

---

[137] *See The Digital Millennium Copyright Act, Section 512*, U.S. Copyright Office, https://www.copyright.gov/dmca/; *Section 512 of Title 17: Resources on Online Service Provider Safe Harbors and Notice-and-Takedown System,* U.S. Copyright Office, https://www.copyright.gov/512/.

[138] Specific methods for compliance with The TAKE IT DOWN Act's notice and removal requirements and the requirements under the TAKE IT DOWN Act are beyond the scope of this article's discussion.

[139] *See, e.g.,* Department of Commerce DMCA Multistakeholder Forum, *DMCA Notice-and-Takedown Processes: List of Good, Bad, and Situational Practices,* https://www.uspto.gov/sites/default/files/documents/DMCA_Good_Bad_and_Situational_Practices_Document-FINAL.pdf.

[140] *See, e.g.,* eBay, Customer Service, *Reporting a product that violates an eBay policy,* https://www.ebay.com/help/policies/member-behavior-policies/reporting-product-violates-ebay-policy?id=4838.

authentic visual depiction, that, when viewed as a whole by a reasonable person, is indistinguishable from an authentic visual depiction of the individual."[141]

As to identifiable individuals in nonconsensual intimate visual depictions, the Act defines "identifiable individual" as an individual: (i) who appears in whole or in part in an intimate visual depiction; and (ii) whose face, likeness, or other distinguishing characteristic (including a unique birthmark or other recognizable feature) is displayed in connection with such intimate visual depiction."[142] It remains to be seen if, as in, *e.g.,* service mark and false association scenarios, items such as tattoos, tattoo arrays,[143] unique costumes or makeup designs associated with specific characters/individuals,[144] or specific contexts associated with an individual[145] might reinforce what would be seen by as reasonable person as contributing to the identification of a specific individual.

For purposes of determining the scope and limitations of a victim's consent – where consent may be a defense – the Act provides that: (A) the fact that the identifiable individual provided consent for the creation of the intimate visual depiction shall not establish that the individual provided consent for the publication of the intimate visual depiction; and (B) the fact that the identifiable individual disclosed the intimate visual depiction to another individual shall not establish that the identifiable individual provided consent for the publication of the intimate visual depiction by the person alleged to have violated the provisions regarding authentic and digitally altered content. Although not specifically limited to "revenge porn," it is clear that the scope of consent provisions are meant to address that species of offense, *i.e.,* where the publisher is reacting to a denial, an end, or an unwanted change in a relationship with the victim. The clarifications arise, at least in part, from typical situations in which there is a voluntary element to the victim's original behavior, but the consent is premised on an expectation of privacy and non-publication.[146] These inferential limitations on consent arise from the seriousness of the injury that can be caused by revenge porn – with the trauma caused by the publication potentially extending to additional traumas of any resultant cyberstalking and harassment by others.[147]

The Act provides, with regard to authentic intimate visual depictions involving adults, that it is unlawful or any person, in interstate or foreign commerce, to use an interactive computer service to knowingly publish an intimate visual depiction of an identifiable individual who is not a

---

[141] *See* Act, Sec. 3, NOTICE AND REMOVAL OF NONCONSENSUAL INTIMATE VISUAL DEPICTIONS, (b)(1)(B), https://www.congress.gov/bill/119th-congress/senate-bill/146/text.
[142] *Id.* at (b)(1)(C).
[143] Davies & Hirji, *Tattoos, Athletes, and Image Rights*, American Bar Association, business law today, Dec. 15,2021, https://businesslawtoday.org/2021/12/tattoos-athletes-and-image-rights/.
[144] *See, e.g.,* Levine, *Pophouse Buys Rights to KISS – Here's What They Have Planned*, billboard, Business News, April 4, 2024, https://www.billboard.com/business/business-news/kiss-rights-pophouse-gene-simmons-1235648622/.
[145] *See, e.g., Hirsch v. S.C. Johnson & Son, Inc.,* 90 Wis. 2d 379 (1979)(football imagery and context reinforced that "Crazy Legs" referred to a specific individual, Elroy Hirsch).
[146] *See, e.g.,* Vile, *Revenge Pornography*, Free Speech Center At Middle Tennessee State University, Updated July 2, 2024, https://firstamendment.mtsu.edu/article/revenge-pornography/.
[147] *Id.*

minor if: (i) the intimate visual depiction was obtained or created under circumstances in which the person knew or reasonably should have known the identifiable individual had a reasonable expectation of privacy; (ii) what is depicted was not voluntarily exposed by the identifiable individual in a public or commercial setting; (iii) what is depicted is not a matter of public concern; and (iv) publication of the intimate visual depiction: (I) is intended to cause harm; or (II) causes harm, including psychological, financial, or reputational harm, to the identifiable individual.[148] Penalties for violation of this Section of the Act include fines under Title 18 of the U.S. Code, imprisonment of not more than two (2) years, or both. In cases where the offense against an adult involves a digital forgery, penalties for violation also include fines under Title 18 of the U.S. Code, imprisonment of not more than two (2) years, or both. Of course, the qualification "what is depicted is not a matter of public concern" is subject to the question: What constitutes a matter of public concern – and at what point (if any) does the otherwise lawful or simply "not illegal" sexual behavior of individuals, including celebrities and politicians, become matters of public concern? It may well be that we need to redefine legitimate "public concern" as something more significant than the satisfaction of the prurient curiosity as the expense of the vulnerable or victimized.[149]

The Act also makes it unlawful for any person to intentionally threaten a violation of the criminal provisions involving authentic intimate visual depictions and an adult victim for the purpose of intimidation, coercion, extortion, or to create mental distress. Penalties for such threats include fines under Title 18 of the U.S. Code, imprisonment of not more than two (2) years, or both. Penalties for threats against adults involving digital forgeries include fines under title 18, United States Code, imprisonment for not more than eighteen (18) months, or both. Arguably, the creation of a digital forgery where there was never any original consent by the victim would merit greater punishment than the use of authentic content. Nonetheless, it appears that the trauma caused by genuine imagery was deemed to be greater than more readily dismissed and discredited digital forgeries. Therefore, there is a disparity in the Act's sentencing provisions.

The Act provides with regard to authentic intimate visual depictions involving minors, *i.e.,* any individual under the age of eighteen (18) years, that it shall be unlawful for any person, in interstate or foreign commerce, to use an interactive computer service to knowingly publish an intimate visual depiction of an identifiable individual who is a minor with intent to: (i) abuse, humiliate, harass, or degrade the minor; or (ii) arouse or gratify the sexual desire of any person. Penalties for violation of this Section of the Act include fines under Title 18 of the U.S. Code, imprisonment of not more than three (3) years, or both. In cases where the offense against a

[148] Act, Sec. 2, CRIMINAL PROHIBITION ON INTENTIONAL DISCLOSURE OF NONCONSENSUAL INTIMATE VISUAL DEPICTIONS,

[149] For those interested in an unusual treatment (or illustration) of this topic, the 2010 documentary, *Tabloid,* is recommended. Although the film is ostensibly about the unusual personality and endeavors of a woman named Joyce McKinney, the behavior of the journalists who were involved – and the presumed "normalcy" they seem to represent – is also an important subject of the film.

minor involves a digital forgery, penalties for violation also include fines under Title 18 of the U.S. Code, imprisonment of not more than three (3) years, or both.

The Act also makes it unlawful for any person to intentionally threaten a violation of the criminal provisions involving authentic intimate visual depictions and a minor victim for the purpose of intimidation, coercion, extortion, or to create mental distress. Penalties for violation of this Section of the Act include fines under Title 18 of the U.S. Code, imprisonment of not more than three (3) years, or both. Penalties for threats against minors involving digital forgeries include fines under title 18, United States Code, imprisonment for not more than thirty (30) months, or both. As with the provisions that apply to violations against adults, there is a disparity in the Act's sentencing provisions with regard to authentic content versus digital forgeries.

Depending on your perspective, it can appear that the sentences in the TAKE IT DOWN Act are too lenient. For example, if the image(s) depict and invite extraordinary abuse that can also intentionally provoke stalking and violence toward the victim from (targeted) unstable viewers, it is hoped that prosecutors will be resourceful in ensuring that charges under additional laws will be brought against the offending individual(s). Frankly, in some jurisdictions, the penalty for stealing a car is a potentially longer sentence (by far) than the terms specified for offenses under the TAKE IT DOWN Act.[150] Perhaps this means those auto theft laws are out of touch, but the point is that the sentences specified under the TAKE IT DOWN Act might seem inordinately short to victims of the violations. Nonetheless, the Act is a step in the right direction – *albeit* in the case of incarceration arguably too short a step – in recognizing the gravity of the situations it addresses.

As provided in the Act, in cases involving the violative publication of intimate visual depictions courts also shall order – in addition to any other sentence imposed and irrespective of any provision of State law – that the convicted person forfeit to the United States: (i) any material distributed in the course of the violation; (ii) the person's interest in property, real or personal, constituting or derived from any gross proceeds of the violation, or any property traceable to such property, obtained or retained directly or indirectly as a result of the violation; and (iii) any personal property of the person used, or intended to be used, in any manner or part, to commit or to facilitate the commission of the violation. The court shall also order restitution for such offenses in the same manner as under section 2264 of title 18, United States Code – which includes the full amount of the victim's losses including: (a) medical services relating to physical, psychiatric, or psychological care; (b) physical and occupational therapy or rehabilitation; (c) necessary transportation, temporary housing, and child care expenses; (d) lost income; (e) attorneys' fees, plus any costs incurred in obtaining a civil protection order; (f) veterinary services relating to physical care for the victim's pet, service animal, emotional

[150] *See, Penalties for Grand Theft Auto*, Justia, https://www.justia.com/criminal/offenses/theft-crimes/auto-theft/.

support animal, or horse; and (g) any other losses suffered by the victim as a proximate result of the offense.[151]

In order to ensure that the Act's criminal provisions do not sweep too broadly, too narrowly (*e.g.,* by interfering with or contradicting other more stringent laws), or become counterproductive, a number of exceptions to the publication sections are provided. In general, the exceptions are illustrative of situations where the requisite malicious intent is absent, the requisite element of intended sexual arousal or gratification is absent, or other laws provide appropriate sanctions. Therefore, the prescriptions and penalties with regard to publications do not apply to:

> (i) a lawfully authorized investigative, protective, or intelligence activity of--
>
> (I) a law enforcement agency of the United States, a State, or a political subdivision of a State; or
>
> (II) an intelligence agency of the United States;
>
> (ii) a disclosure made reasonably and in good faith--
>
> (I) to a law enforcement officer or agency;
>
> (II) as part of a document production or filing associated with a legal proceeding;
>
> (III) as part of medical education, diagnosis, or treatment or for a legitimate medical, scientific, or education purpose;
>
> (IV) in the reporting of unlawful

[151] *See* 18 U.S. Code § 2264 – Restitution, https://www.law.cornell.edu/uscode/text/18/2264.

content or unsolicited or unwelcome conduct or in pursuance of a legal, professional, or other lawful obligation; or

(V) to seek support or help with respect to the receipt of an unsolicited intimate visual depiction;

(iii) a disclosure reasonably intended to assist the identifiable individual;

(iv) a person who possesses or publishes a digital forgery of himself or herself engaged in nudity or sexually explicit conduct (as that term is defined in section 2256(2)(A) of title 18, United States Code); or

(v) the publication of an intimate visual depiction that constitutes--

(I) child pornography (as that term is defined in section 2256 of title 18, United States Code); or

(II) a visual depiction described in subsection (a) or (b) of section 1466A of title 18, United States Code (relating to obscene visual representations of the sexual abuse of children).[152]

[152] *See* Act, Sec. 2, CRIMINAL PROHIBITION ON INTENTIONAL DISCLOSURE OF NONCONSENSUAL INTIMATE VISUAL DEPICTIONS, (2)(C) , https://www.congress.gov/bill/119th-congress/senate-bill/146/text.

The exceptions are sure to raise a number of issues, and at least one organization has noted its "objection to the exception provided for 'a person who possesses or publishes an intimate visual depiction of himself or herself,' which creates a dangerous loophole that would seemingly allow a person to disclose intimate images without consent so long as that person also appears in the image."[153] Although this interpretation of the exception is subject to debate – *e.*g., the exception arguably relates to only to an individual's depictions of himself or herself, with the addition of another party providing a basis for prosecution with regard to that other individual – there can be no doubt that defense attorneys will seek to impose the most expansive interpretations of the exceptions, unless a narrow interpretation help to attack the general legitimacy of the criminal provisions.

With regard to challenges to the Act, Legislative Attorney Victoria L. Killion offers excellent insights (current as of 05/20/25) into the Act and potential challenges, including the following:

> Because the TAKE IT DOWN Act regulates speech in the form of visual depictions, First Amendment questions may be at the forefront of any legal challenges. In particular, because the Act regulates speech on the basis of its content (*i.e.,* sexually explicit depictions), it could receive the most rigorous form of First Amendment scrutiny (strict scrutiny) if challenged in court. Under that standard, the government would need to show that the law is the least restrictive means of advancing a compelling governmental interest. In the event of a free-speech challenge to the Act's criminal prohibitions, a reviewing court might consider judicial decisions resolving First Amendment challenges to similar state nonconsensual pornography laws. The highest courts of six states (IL, IN, MN, NE, TX, and VT) have upheld their states' nonconsensual pornography laws against free-speech challenges, though one court did so in a nonprecedential (*i.e.,* nonbinding) opinion. (The criminal prohibitions in the TAKE IT DOWN Act share some of the features of the state laws upheld in those jurisdictions (most of which applied strict scrutiny), including rigorous mental state requirements (knowingly or specific intent), a focus on depictions of identifiable individuals, and exceptions for circumstances where a depicted adult does not have a reasonable expectation of privacy or the depiction is a matter of public concern (links omitted.)[154]

Finally, as noted in recent commentary by the National Association of Attorney's General, "[t]he Take It Down Act does not include a preemption clause, allowing state nonconsensual intimate

---

[153] Franks, *CCRI Statement on the Passage of the TAKE IT DOWN Act (S. 146)*, Cyber Civil Rights Initiative, April 28, 2025, https://cybercivilrights.org/ccri-statement-on-the-passage-of-the-take-it-down-act-s-146/.

[154] Killion, *The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images*, Congress.Gov, Congressional Research Service, https://www.congress.gov/crs-product/LSB11314. *See also*, Mullin, *The TAKE IT DOWN Act: A Flawed Attempt to Protect Victims That Will Lead to Censorship*, Electronic Frontier Foundation, Feb. 11, 2025 for a critical view of some aspects of the Act, https://www.eff.org/deeplinks/2025/02/take-it-down-act-flawed-attempt-protect-victims-will-lead-censorship.

imagery (NCII) laws to work concurrently.[155] This is important because all fifty states, the District of Columbia, Puerto Rico, and Guam have laws against nonconsensual distribution of intimate images.[156] However, the TAKE IT DOWN Act adds a degree of federal involvement and expressly goes beyond unmodified NCII – adding, unlike some local laws, proscriptions against AI-assisted "digital forgeries."

By now, this question might occur to some readers: If we're talking about deepfakes and standards for assessment and use of AI-assisted deepfake evidence, why does the AI Action Plan and this article devote time to the TAKE IT DOWN Act and other generative AI issues? Fair question. Here's a proposed answer:

VI. CONCLUSION FOR PARTS 2 & 3

The AI Action Plan's section on Combating Synthetic Media in the Legal System highlights the TAKE IT DOWN Act because the Act is premised, in part, on a recognition of, and intense focus on, the immediate dangers of deepfake technology – and also shows the benefits of "across-the-aisle" political and community action to combat the very real and very present threats of injurious uses of deepfake outputs. Frankly, delay and endless debate were not options. The same holds true with regard to addressing the generation and misuse of deepfake evidence in our judicial and other evidence-driven deliberative processes. Rules without practical means to apply, support, and sustain them will simply fall short if not backed by resources, such as the program noted in the pertinent section's First Policy Recommendation and the activism and sharing of expertise contemplated in the Second and Third Policy Recommendations. Somewhere, someone might be contemplating calling their deepfake identification and refutation efforts "The Putnam-Mather Project." That's a little indelicate, but it has the advantage of avoiding some more contemporary – and potentially divisive – examples of outrageously concocted evidence that has recently entered our courtrooms and deliberative processes. As with the rapid processes that led to the formation and passage of the TAKE IT DOWN Act, we need a rapid and coherent development of deepfake detection resources that can be made readily available in criminal, civil, and other evidence-driven processes that can affect us individually and can determine the nature of our government and society. Again, Rules without generally-available resources to properly administer them – including, for example, heightened federal development and sharing of deepfake detection resources under the AI Action Plan – leave the job only half-finished. Now, let's talk in more detail about LLM-related dangers and the invasion of the libertine chatbots. (To be continued . . .)

---

[155] Leibert, *Congress's Attempt to Criminalize Nonconsensual Intimate Imagery: The Benefits and Potential Shortcomings of the TAKE IT DOWN Act*, National Association of Attorney's General, Aug. 26, 2025, https://www.naag.org/attorney-general-journal/congresss-attempt-to-criminalize-nonconsensual-intimate-imagery-the-benefits-and-potential-shortcomings-of-the-take-it-down-act/.

[156] *See Nonconsensual Distribution of Intimate Images*, Cyber Civil Rights Initiative, (visited September 2, 2025, https://cybercivilrights.org/nonconsensual-distribution-of-intimate-images/.