THE COLUMBIA JOURNAL OF
LAW & the ARTS

A QUARTERLY JOURNAL OF LAW AND THE ARTS,
ENTERTAINMENT, COMMUNICATIONS, AND INTELLECTUAL PROPERTY

2025 Symposium of the Kernochan Center for Law, Media and the Arts:
"Deepfakes: In Search of Global Solutions"

Vol. 49, No. 4 ♦ Symposium Issue

COLUMBIA LAW SCHOOL

THE COLUMBIA JOURNAL OF
LAW & the ARTS

VOL. 49 **NO. 4**

CONTENTS

Produced and published by members of the Columbia Journal of Law & the Arts.
All members are students at Columbia Law School.

Columbia Journal of Law & the Arts (JLA)
Columbia Law School
435 West 116th Street
New York, NY 10027
jla-editor@law.columbia.edu
http://www.lawandarts.org

ISSN (Print): 1544-4848
ISSN (Online): 2161-9271

Cite this volume: 49 COLUM. J.L. & ARTS.

Typeset in Crimson Text.

Printed in the United States of America.

Printed by Sheridan PA, a CJK Group Company

Sheridan
450 Fame Avenue
Hanover, PA 17331

THE COLUMBIA JOURNAL OF
LAW & the ARTS

VOL. 49 **NO. 4**

VOL. 49 MASTHEAD (2025–2026)

Editor's Note on the Symposium Issue

Each academic year, the *Columbia Journal of Law & the Arts* publishes an Issue dedicated to the annual Symposium of the Kernochan Center for Law, Media and the Arts, which is hosted at Columbia Law School. This year's Symposium was titled Deepfakes: In Search of Global Solutions and was held on Friday, October 24, 2025. As always, the Journal was honored to participate in the event and is pleased to publish the proceedings here.

There are two types of publications in this Issue. Each speaker was asked to select one of the two options: to write an Article based on his or her remarks at the Symposium or to produce a Transcript of his or her remarks. The Articles have been written, edited, and proofread to the same high standard as other academic articles published by the *Journal* in its non-Symposium Issues. The Transcripts have been edited lightly for concision and clarity. The pieces in this Issue are presented in the order in which contributors spoke at the Symposium. The Program of the 2025 Symposium on page vii of this Issue reflects the actual order of the speakers on the day of the event.

More information about the 2025 Symposium can be found on the Kernochan Center's website,[1] including readings for the event,[2] biographies of the speakers,[3] and video recording of the event.[4]

1. The 2025 Symposium website is at https://kernochan.law.columbia.edu/content/symposium-2025-deepfakes [/web/20260113141606/https://kernochan.law.columbia.edu/content/symposium-2025-deepfakes].

2. *See supra* note 1 for readings.

3. *See supra* note 1 for speaker biographies.

4. Video recording of the 2025 Symposium can be found at https://kernochan.law.columbia.edu/content/kernochan-symposium-2025-part-1 [/web/20260113141913/https://kernochan.law.columbia.edu/content/kernochan-symposium-2025-part-1] and https://kernochan.law.columbia.edu/content/kernochan-symposium-2025-part-2 [/web/20260113143338/https://kernochan.law.columbia.edu/content/kernochan-symposium-2025-part-2].

Program of the 2025 Symposium

SESSION I—THE EXISTENCE (OR NOT) OF RIGHTS AGAINST UNAUTHORIZED DEEPFAKES

A. PANEL I—DEMONSTRATION OF AI IMAGE-GENERATION AND MUSICAL PERFORMANCE DEEPFAKES

Moderator:
Philippa Loengard

Demonstration:
Makena Binker Cosen

Comment:
Jane Ginsburg

B. PANEL II—SCOPE OF INDIVIDUAL RIGHTS: A COMPARATIVE ANALYSIS

Moderator:
Jane Ginsburg

Speakers:
Jennifer E. Rothman
Ben S. Sheffner
Valérie-Laure Benabou
Graeme W. Austin

C. PANEL III—PUBLIC RIGHTS: TRANSPARENCY OBLIGATIONS

Moderator:
Philippa Loengard

Speaker:
Célia Zolynski

Comment:
Olivier Sylvain

SESSION II—ENFORCEMENT

A. PANEL I—JURISDICTION AND APPLICABLE LAW

Moderator:
Jane Ginsburg

Speakers:
Edouard Treppoz

Comment:
Graeme B. Dinwoodie

B. PANEL II—AI DEVELOPERS' AND PLATFORMS' DERIVATIVE LIABILITY

Moderator:
Caitlin McGrail

Speaker:
Shyamkrishna Balganesh

Comments:
Edouard Treppoz
Valérie-Laure Benabou
Graeme W. Austin

C. PANEL III—CIVIL ENFORCEMENT

Moderator:
Philippa Loengard

Speaker:
David S. Louk

Comments:
Célia Zolynski
Edouard Treppoz

Illusions of Control

Consumer Digital Replica Platforms and the Governance of Identity

Makena Binker Cosen*

* Makena Binker Cosen is an Associate at Kirkland & Ellis LLP in New York. She graduated Columbia Law School in 2025.

To me, the Symposium has become a marker in the passage of time. I began attending before I was even sure that law school was for me, and the enthusiasm in the room played a meaningful role in that decision. Each year, I returned with a clearer grasp of the conversations unfolding around me—eventually participating in discussions I once worked so hard just to follow. As I became familiar with the faces reappearing year after year, the Symposium has come to feel like a reunion of a community I am honored to be part of.

Returning after graduating law school as a speaker, standing alongside the mentors who had shaped my path, carried real personal significance. That is why, when my phone rang in the middle of my presentation on deepfakes, I should have been mortified. Instead, when I answered the call, I was confident knowing that my digital replica would step in for me. Deliberately planned, this moment illustrated a future in which the same technologies often associated with deepfakes are used to create digital replicas that make casual appearances in everyday life. Built with amateur knowledge, on a tight budget, and in a matter of days, consumer digital replica platforms are already ushering in that future.

Digital replicas are becoming ordinary. People record short samples of speech to preserve a voice, upload photographs to generate images they never posed for, or rely on systems trained on their prior writing to draft messages on their behalf. These practices are rarely framed as moments of identity delegation. They are presented instead as conveniences, accessibility tools, or creative efficiencies. Yet each depends on the same underlying move: extracting elements of a real person's identity and embedding them into a persistent digital system capable of acting without the person's ongoing participation. Once created, these systems can speak, appear, or respond in ways meaningfully attributable to an individual, even as control over their operation increasingly rests with the platform that hosts them. Existing legal frameworks have struggled to account for that shift.

This Article argues that consumer digital replicas constitute a distinct category of identity-bearing systems that cannot be adequately understood through doctrines

focused on deception, labor, or isolated misuse. The central problem is not that digital replicas are inherently harmful. Many are beneficial and empowering. The problem is governance: how identity becomes persistent, generative, and platform-mediated over time.

Part I stabilizes "digital replica" as an analytical category. It traces the term's reactive evolution across preservation, engineering, entertainment, labor, and AI governance, and distills shared structural features that cut across those contexts. From this synthesis, the Comment proposes a working definition suited to the consumer era. Part II examines how digital replicas enter everyday life through consumer products. Focusing on voice, visual, and behavioral systems, it shows how modest user inputs generate reusable identity-signifying outputs, even as these tools are framed as neutral utilities rather than identity technologies. Part III turns to governance. It analyzes the limits of revocable consent and the ongoing contest over permissible use shaped by platform design, safeguards, pricing, and portability. Across the lifecycle of a replica, control migrates incrementally from users to platforms. The Article concludes by explaining why existing legal frameworks arrive too late. Without a structural account of digital replicas as persistent identity systems, law intervenes only after control over identity has already been decided.

I. DEFINING DIGITAL REPLICAS

The term "digital replica" has become a familiar part of contemporary debates about artificial intelligence, identity, and personality rights. Yet despite its frequent use, the term does not carry a single, uniform meaning across legal, technical, and cultural contexts. It has been used to describe everything from digital museum objects to CGI "digital doubles" of actors, AI-generated voice models, and synthetic media regulated under deepfake statutes. These uses did not develop together, and they do not rest on a shared understanding of what, exactly, is being replicated—or why it matters.

For legal purposes, this definitional fragmentation is a problem. Treating digital replicas as merely deceptive deepfakes or as a niche labor issue misses a broader and increasingly common phenomenon: the creation of persistent digital systems that reproduce aspects of a real human identity and can operate independently of the person they represent. This section will develop a definition of "digital replica" by examining the term's evolution, isolating its defining features, and distinguishing it from adjacent concepts.

A. THE REACTIVE EVOLUTION OF THE TERM "DIGITAL REPLICA"

The phrase "digital replica" did not begin as a legal concept, nor did it initially refer to people at all. Its earliest uses appeared in technical and academic contexts concerned with faithfully reproducing physical objects in digital form. In cultural heritage scholarship, for example, a digital replica was defined as "a faithful copy of an original artifact in the digital domain, including its appearance, its morphology and how it is

meant to interact."[1] In that setting, replication was about accuracy and preservation. The object being replicated was static, nonhuman, and socially inert. Legal questions about consent, control, or misuse were simply not part of the picture.

A more dynamic version of the term emerged in engineering and industrial systems design, where "digital replica" became closely associated with the idea of a "digital twin." Industry sources described digital replicas as virtual representations of physical assets, systems, or processes, often connected to real-time data and used for monitoring, simulation, or optimization.[2] McKinsey, for instance, defined a digital twin as "a digital replica of a physical object, person, system, or process, contextualized in a digital version of its environment."[3] These definitions introduced persistence and reusability as core features, but only in service of operational goals. Even when people were included as possible subjects, they were treated as systems to be modeled, not as identities to be protected.

The term took on a very different meaning once advances in computer-generated imagery (CGI) and machine learning made it possible to replicate human likeness. By the late 2010s, media outlets began describing "digital replicas" of actors as highly realistic digital doubles capable of performing on screen without the actor's physical presence—including, in some cases, after the actor's death.[4] Legal scholarship soon followed. In the entertainment context, a digital replica was defined as "a computer-generated image that recreates the likeness of a person—their face, body, voice, and movement."[5] At this point, replication was no longer about copying objects or optimizing systems. It was about standing in for a human being.

Labor organizations adopted the term with similar assumptions. SAG-AFTRA defined a "digital replica" as "a replica of your voice and/or likeness that is created using digital technology, such as artificial intelligence," and drew a sharp distinction between replicas created with a performer's participation and those created independently from existing recordings.[6] These definitions were narrow by design. They focused on

1. Fabrizio Ivan Apollonio, Marco Gaiani & Simone Garagnani, *Visualization and Fruition of Cultural Heritage in the Knowledge-Intensive Society*, *in* HANDBOOK OF RESEARCH ON IMPLEMENTING DIGITAL REALITY AND INTERACTIVE TECHNOLOGIES TO ACHIEVE SOCIETY 5.0 471, 495 (Francesca Maria Ugliotti & Anna Osello eds., 2022).

2. *What Is Digital-Twin Technology?*, MCKINSEY & CO. (Aug. 26, 2024), https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-digital-twin-technology [https://web.archive.org/web/20260205210906/https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-digital-twin-technology].

3. *Id.*

4. Sean Cummings, *Q&A: Stanford Engineers Discuss Digital Doubles*, STAN. REP. (Sep. 26, 2023), https://news.stanford.edu/stories/2023/09/qa-stanford-engineers-discuss-digital-doubles [https://web.archive.org/web/20250818120434/https://news.stanford.edu/stories/2023/09/qa-stanford-engineers-discuss-digital-doubles].

5. Alexandra Curren, Note, *Digital Replicas: Harm Caused by Actors' Digital Twins and Hope Provided by the Right of Publicity*, 102 TEX. L. REV. 155, 159 (2023).

6. *Digital Replicas 101*, SAG-AFTRA 1 (2023), https://www.sagaftra.org/sites/default/files/sa_documents/DigitalReplicas.pdf [https://web.archive.org/web/20241003065947/https://www.sagaftra.org/sites/default/files/sa_documents/DigitalReplicas.pdf].

professional performers, third-party creators, and unauthorized reuse. The problem was framed as one of bargaining power and consent in employment relationships—not as a general issue of identity governance.

Legislatures then picked up the term in response to generative AI and deepfakes, narrowing it further. The European Union's Artificial Intelligence Act – hailed as the first comprehensive AI framework of its kind—did not define the concept of a "digital replica." However, it addressed deep fakes as including AI-generated media that "resembles existing persons" and "would falsely appear to a person to be authentic."[7] By contrast, U.S. state laws expressly adopt the language of "digital replicas." For example, New York defines a digital replica as "a digital simulation of the voice or likeness of an individual that so closely resembles the individual's voice or likeness that a layperson would not be able to readily distinguish the digital simulation from the individual's authentic voice or likeness"; California and Illinois likewise emphasize high realism and the likelihood of audience confusion.[8] Federal policy echoes this framing: The U.S. Copyright Office has described digital replicas as digital recordings that "realistically but falsely depict an individual."[9]

These definitions make sense given their regulatory goals. They are aimed at fraud, impersonation, and misinformation, and they therefore focus on realism and deception. However, they also reflect assumptions that are increasingly misaligned with the direction in which digital replica technologies are evolving. While such systems are not yet mainstream, they are becoming progressively more accessible, and are often designed for disclosed, consented, and self-directed use rather than deception or impersonation.

Consumer and accessibility technologies illustrate this shift. Voice banking tools and personal AI systems allow users to create persistent models trained on their own speech or expressive patterns. Apple's Personal Voice feature, for example, enables users to generate a synthetic voice derived from their own recordings for ongoing use.[10]

---

7. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, Laying Down Harmonised Rules on Artificial Intelligence, art. 3(60), 2024 O.J. (L 1689) 1; Shiona McCallum, Liv McMahon & Tom Singleton, *MEPs Approve World's First Comprehensive AI Law*, BBC (Mar. 13, 2024), https://www.bbc.com/news/technology-68546450 [https://web.archive.org/web/20260226030228/https://www.bbc.com/news/technology-68546450].

8. N.Y. GEN. OBLIG. LAW § 5-302 (McKinney 2025); CAL. LAB. CODE § 927(c)(1) (West 2025); 765 ILL. COMP. STAT. 1075/5 (West 2025); *see also* Miguel A. Lopez, Brad Kelley & Shreya Mantrala, *New York's Digital Replica Law and Its Impact on Artificial Intelligence and the Entertainment Industry*, LITTLER (Jan. 21, 2025), https://www.littler.com/publication-press/publication/new-yorks-digital-replica-law-and-its-impact-artificial-intelligence [https://web.archive.org/web/20260213181856/https://www.littler.com/news-analysis/asap/new-yorks-digital-replica-law-and-its-impact-artificial-intelligence-and].

9. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE, PART I: DIGITAL REPLICAS 2 (2024), https://copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf [https://web.archive.org/web/20260205212401/https://copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf].

10. Apple, *Apple Previews Live Speech, Personal Voice, and More New Accessibility Features* (May 16, 2023), https://www.apple.com/ke/newsroom/2023/05/apple-previews-live-speech-personal-voice-and-more-new-accessibility-features

These systems are rarely called "digital replicas," but they share the same underlying features found in other contexts: persistence, reusability, and the ability to generate identity-signifying outputs without live human performance.[11]

Looking across these domains, a pattern becomes clear. "Digital replica" has been defined reactively, with each definition tailored to a specific institutional concern—preservation, optimization, labor protection, or deception. None of these definitions are wrong, but each is incomplete. Together, they obscure a broader trend that cuts across contexts: the externalization of human identity into persistent digital systems that can be stored, modified, licensed, and reused over time.

B. CORE CHARACTERISTICS OF DIGITAL REPLICAS

A more general definition of digital replicas must therefore move away from sector-specific use cases and toward shared structural features. What unites the disparate technologies described above is not how they are deployed, but how they function. These features also clarify why certain adjacent technologies—deepfakes, avatars, filters, and generic or fictional AI systems—fall outside the category, even where they overlap in form or technical sophistication.

First, a digital replica is defined by referential fidelity to a real person. It is tied to an identifiable individual and derives its meaning from that connection, whether the fidelity is visual, vocal, behavioral, or expressive. Perfect realism is not required. What matters is that the system purports to stand in for a particular person, not the medium used or whether the output is deceptive. By contrast, generic AI outputs, fictional characters, and virtual influencers lack a real-world referent and therefore fall outside the category.

Second, digital replicas exhibit technical persistence and reusability. Unlike a single photograph, recording, or live performance, a digital replica is designed to endure: It can be stored, copied, updated, and redeployed across contexts, often without the person's ongoing involvement.[12] This distinguishes replicas from one-off deepfakes

[https://web.archive.org/web/20260205212356/https://www.apple.com/ke/newsroom/2023/05/apple-previews-live-speech-personal-voice-and-more-new-accessibility-features/].

11. Consumer discourse reflects an understanding of these features, with one technology commentator seeking to re-deploy Apple's Personal Voice to screen calls from telemarketers, generate narrations for videos and presentations, and enhance live speech in professional settings on sick days. *See* Brad Morton, *Five Ways I Wish Apple Would Let Me Use My AI Cloned Voice*, HOW-TO GEEK (June 3, 2024), https://www.howtogeek.com/5-ways-i-wish-apple-would-let-me-use-my-ai-cloned-voice/ [https://web.archive.org/web/20260326155141/https://www.howtogeek.com/5-ways-i-wish-apple-would-let-me-use-my-ai-cloned-voice/].

12. *See, e.g.*, *Auto-Updating Knowledge Base*, PICKAXE, https://pickaxe.co/ai-knowledge-base [https://web.archive.org/web/20260326155455/https://pickaxe.co/ai-knowledge-base] (last visited Mar. 26, 2026) (automatically updates AI agent's sources on a daily basis "without any manual work"); *Use Cases: Unlock New Ways to Create with HeyGen*, HEYGEN, https://www.heygen.com/use-cases [https://web.archive.org/web/20260326155641/https://www.heygen.com/use-cases] (last visited Mar. 26, 2026) (showing AI avatar can be used for social media, advertising, training, presentations, and courses); Sydney Bradley, *Death Isn't the End: Meta Patented an AI That Lets You Keep Posting from Beyond the Grave,* BUS. INSIDER (Feb. 11, 2026), https://www.businessinsider.com/meta-granted-patent-for-ai-llm-bot-dead-

and real-time filters, which remain tied to a single use or moment. Persistence matters legally because it allows identity representations to accumulate meaning over time, appear in unanticipated settings, and outlast changes in the individual's preferences—or even their life.

Third, digital replicas involve a separation of identity from contemporaneous human action. Traditionally, speech, performance, and appearance are inseparable from the person producing them in real time. Digital replicas break that link. They allow identity-signifying outputs to be generated autonomously or semi-autonomously-speaking without the person speaking, performing without the person performing.[13] This feature distinguishes replicas from avatars and similar tools that mediate live user action, rather than functioning as independent stand-ins.

Finally, modern digital replicas are capable of recombination and automation. Rather than merely replaying a stored likeness, they apply learned identity traits across new prompts and contexts to generate novel outputs over time.[14] A replica may speak on topics the individual never addressed, respond to questions never anticipated, or combine elements of style, tone, and content the individual never authored. This generative capacity explains why neither realism nor deception alone defines the category. Legally, the significance lies in derivation: a limited act of creation can produce an open-ended stream of identity-signifying outputs, transforming the replica from a finite representation into an ongoing source of identity.

C. A WORKING DEFINITION FOR CONSUMER-ERA DIGITAL REPLICATION

Taken together, these features support the following working definition:

> A digital replica is a persistent, reusable digital system that is referentially anchored to a specific, identifiable human being and capable of generating identity-signifying outputs—such as voice, appearance, behavior, or expressive style—without contemporaneous human performance.

This definition is deliberately forward-looking. It does not hinge on deception, perfect realism, or malicious intent. Instead, it captures the structural features that give rise to recurring legal and governance concerns across consumer, enterprise, and creative contexts.

---

paused-accounts-2026-2 [https://web.archive.org/web/20260326155909/https://www.businessinsider.com/meta-granted-patent-for-ai-llm-bot-dead-paused-accounts-2026-2].

13. *See* Pietro Ruiu et al., *Metaverse & Human Digital Twin: Digital Identity, Biometrics, and Privacy in the Future Virtual Worlds*, 8 MULTIMODAL TECH. & INTERACTION 48, 50–51 (2024).

14. Ajay Bandi, Pydi Venkata Satya Ramesh Adapa & Yudu Eswar Vinay Pratap Kumar Kuchi, *The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges*, 15 FUTURE INTERNET 260, 268–69 (2023).

II. CONSUMER-FACING DIGITAL REPLICA PRODUCTS

Digital replicas enter everyday life not as a single, unified technology, but through a growing ecosystem of consumer tools that repurpose complex generative systems for familiar tasks. Users are invited to record speech, upload images or video, or provide prior writings in order to generate new content that resembles their own voice, appearance, or manner of response. What distinguishes these products is not merely what they produce, but how they are framed: as tools for accessibility or to facilitate content creation, communication, or continuity.[15] Across voice, visual, and behavioral domains, increasingly sophisticated systems are introduced through interfaces that emphasize ease and speed, while the technical processes that allow identity-related traits to persist and recombine remain largely invisible.

A. VOICE REPLICATION

Voice replication is often the first point at which consumers encounter digital replica technologies, in part because it builds directly on familiar practices.[16] The basic user action is simple: recording speech or uploading existing audio. This can involve reading a short script, providing a few minutes of prior recordings, or recording sample phrases through a web interface. The output is a synthetic voice capable of generating new speech from text, without requiring the user to re-record every time.

In practice, voice replicas are used in a wide range of everyday applications. Creators use them to generate voiceovers for videos or podcasts without repeated recording sessions.[17] Speakers and presenters use them to correct mistakes in recorded talks by typing revised sentences rather than re-recording entire segments. Media producers use voice replicas to dub content into other languages while preserving a familiar vocal identity. For users with speech impairments or degenerative conditions, voice replication can support accessibility by recreating a person's own voice for assistive communication.

Technically, these systems rely on prebuilt speech models that already capture how human voices work in general. The user's recordings are used to adapt that system to sound like a particular person, either almost instantly or through a longer refinement process. From the user's perspective, however, the mechanics are abstracted away. The experience is one of typing text and receiving speech. The resulting voice is not a fixed

15. *See ElevenLabs Impact Program*, ELEVENLABS, https://elevenlabs.io/impact-program [https://web.archive.org/web/20260326160546/https://elevenlabs.io/impact-program] (last visited Mar. 26, 2026); *Voice Cloning: Create a Replica of Your Voice That Sounds Just Like You*, ELEVENLABS, https://elevenlabs.io/voice-cloning [https://web.archive.org/web/20260326160506/https://elevenlabs.io/voice-cloning] (last visited Mar. 26, 2026); Bradley, *supra* note 12.

16. For example, ElevenLabs asks users to upload or record ten to thirty seconds of audio to generate a voice clone. *Voice Cloning*, *supra* note 15.

17. *Id.*

recording, but a reusable system that can speak new words, in new contexts, long after the initial input.[18]

Voice replication also functions as a foundational layer for other forms of digital replication. Many visual replica systems—particularly talking-head videos and avatar presenters—combine synthetic video with synthetic voice, integrating voice models directly into visual pipelines.

B. VISUAL REPLICATION

Visual replication encompasses a broader and more varied set of consumer products than video alone. At the simplest level, users upload a set of photographs—often taken from different angles or with different expressions—to generate new images placing them into novel scenarios.[19] Popular applications include generating portraits in imagined roles or settings, such as historical figures, fantasy characters, or alternate professions. These systems replicate not only facial features but often body shape, posture, clothing, and overall presentation, producing images that depict the user in contexts they never physically occupied.

More advanced workflows build on this foundation. Filmmakers and creators increasingly begin with generative still images—concept art, storyboards, or character portraits—and then use AI tools to animate those images, adding motion, facial expression, and lip synchronization.[20] Earlier consumer tools required a progression from still images to animated portraits to short video clips. Improvements in generative video models are collapsing these steps, making it increasingly possible to generate video directly rather than animating a static image after the fact.[21]

User input scales with output capability. Uploading a single image supports limited animation; providing multiple images or minutes of video footage allows the system to learn how a face and body move over time.[22] Some platforms guide users through

18. *See* Genesis Gregorious Genelza, *A Systematic Literature Review on AI Voice Cloning Generator: A Game Changer or a Threat?*, *4 J. EMERGING TECHS.* 54, 55 (2024).

19. *See Most Realistic AI Image-to-Video Animation Tools*, AKOOL (July 11, 2025), https://akool.com/blog-posts/best-5-ai-video-generators-free-for-turning-images-into-video [https://web.archive.org/web/20260205213915/https://akool.com/blog-posts/best-5-ai-video-generators-free-for-turning-images-into-videos].

20. *See id.*

21. *See, e.g.*, *AI Video Generator*, ADOBE, https://www.adobe.com/products/firefly/features/ai-video-generator.html [https://perma.cc/EA2C-KA4N?type=image] (last visited Mar. 26, 2026); *Text-to-Video: A Step-by-Step Guide to Creating Your First Video Using AI*, KLING AI (Aug. 29, 2025), https://kling.ai/blog/text-to-video-ai-creation-guide [https://perma.cc/MF3U-7RSB]; *Free AI Video Creation*, OPENART, https://openart.ai/suite/create-video/kling-3-omni [https://perma.cc/TYP8-YT8X] (last visited Mar. 26, 2026).

22. *See, e.g.*, David Ephraim, *The Complete Guide to Creating AI Avatars: From FLUX.1 to Video Generation*, ATAK INTERACTIVE (Aug. 1, 2025), https://www.atakinteractive.com/blog/the-complete-guide-to-creating-ai-avatars-from-flux.1-to-video-generation [https://web.archive.org/web/20260326162045/https://www.atakinteractive.com/blog/the-complete-guide-to-creating-ai-avatars-from-flux.1-to-video-generation] (suggesting using fifteen to fifty photos or at least ten videos, each ten seconds long).

capture sessions, asking them to read scripted sentences aloud to a camera so the system can model facial motion during speech. The result can be a reusable digital presenter capable of delivering new scripts on demand.

These visual replicas are widely used for content creation and professional communication. Companies deploy synthetic talking heads for corporate training videos, onboarding materials, and internal communications.[23] Marketers use them to generate personalized or localized video content at scale. Educators and nonprofits use animated historical figures or spokespersons to deliver information in engaging formats. In many of these cases, visual replicas are tightly integrated with voice replication technologies, producing a combined audiovisual stand-in that can be updated without re-recording.

As visual replicas have become more capable, the systems that generate them have also become more mediated. What once required studio-grade CGI, motion capture, or manual compositing can now be accomplished through browser-based tools. At the same time, the rendering, animation, and storage processes are handled entirely by platforms, leaving users with clear inputs and outputs but little visibility into how their likeness is maintained or extended over time.

C. BEHAVIORAL AND PERSONA REPLICATION

Behavioral and persona replication systems operate through a different form of user contribution. Rather than capturing new media, these tools are built from what users have already produced: emails, documents, messages, posts, or transcripts.[24] The system analyzes these materials to learn patterns not only of language, but often of preferences, expertise, or values reflected in prior decisions and responses.

The resulting replicas are used in a variety of roles. Some are designed to draft or respond to communications in a user's typical style. Others function as customer-support or fan-engagement chatbots, answering questions in a way that reflects a particular brand voice, creator persona, or subject-matter expertise.[25] More personal applications include companionship agents or griefbots trained on the communications

23. *See* Roberto Gozalo-Brizuela & Eduardo Garrido-Merchán, *Applications of Video Generation Models in the Journal of Computer Science*, 20 J. COMPUT. SCI. 801, 801–818 (2024).

24. *See How It Works*, TONECLONE https://toneclone.ai/#how-it-works [https://web.archive.org/web/20260326162428/https://toneclone.ai/#how-it-works] (last visited Mar. 26. 2026); Brooke Steinberg, *People Are Making Digital Clones" of Themselves—To Do Their Work for Them*, N.Y. POST (Apr. 15, 2024), https://nypost.com/2024/04/15/lifestyle/people-making-digital-clones-of-themselves-to-do-their-work/ [https://web.archive.org/web/20260216065115/https://nypost.com/2024/04/15/lifestyle/people-making-digital-clones-of-themselves-to-do-their-work/]; Amanda Caswell, *I Made a Digital Twin of Myself in ChatGPT—And It Changed How I Work Every Day*, TOM'S GUIDE (Feb. 17, 2026) https://www.tomsguide.com/ai/i-made-a-digital-twin-of-myself-in-chatgpt-and-it-changed-how-i-work-every-day [https://web.archive.org/web/20260308222028/https://www.tomsguide.com/ai/i-made-a-digital-twin-of-myself-in-chatgpt-and-it-changed-how-i-work-every-day].

25. *See* Heng Gu, Senthil Chandrasegaran & Peter Lloyd, *Synthetic Users: Insights From Designers' Interactions With Persona-Based Chatbots*, 39 A.I. FOR ENG'R, DESIGN, ANALYSIS & MFG. 1, 3 (2025).

of deceased individuals, allowing family members or loved ones to interact with a simulation shaped by that person's prior expressions.

In these systems, the output is not a static artifact but an ongoing interaction. Users prompt the system with questions or scenarios, and the replica generates responses dynamically. Unlike voice or visual replicas, behavioral systems are often constrained not just by how someone speaks, but by what they are understood to know, believe, or prioritize. A replica trained on professional correspondence may answer technical questions; one trained on personal messages may reproduce emotional tone or relational patterns.

From the user's perspective, setup may involve uploading archives, linking accounts, or interacting with the system over time. Once established, the system operates with a degree of autonomy, generating responses that feel personally attributable even as the mechanisms that produce them remain opaque.[26]

D. PATTERNS ACROSS CONSUMER SYSTEMS

Across voice, visual, and behavioral domains, consumer digital replicas share a common structure: modest user inputs yield systems capable of producing new outputs across time and context. As these tools have matured, two features have driven their widespread adoption. First, refinement: Outputs have become more natural, flexible, and convincing. Second, accessibility: Platforms increasingly make these capabilities cheap, fast, and easy to use. Just as importantly, they are introduced and experienced as tools—for editing, creating, communicating, or scaling—not as technologies of identity.

That framing has proven durable even as capabilities expand. Voice and visual replication, once distinct, are already routinely combined in talking-head videos, avatar presenters, and synthetic media workflows. The likely next step is not simply better quality, but broader integration: the addition of behavioral replication to systems that already speak and appear. At that point, digital replicas would not only sound and look like a person, but also respond, prioritize, or interact in ways associated with that person.

Whether such systems will continue to be understood as tools or begin to feel like extensions of the self remains an open question. What is clear is that this evolution is unfolding against a backdrop of normalization. By the time more fully integrated replicas become feasible, the practices that enable them—recording, uploading, delegating—will already feel routine. That context matters. It shapes not only how these technologies are adopted, but how their significance is recognized, or overlooked, as they move from helpful utilities toward something closer to a stand-in.

26. *See, e.g.*, PICKAXE, *supra* note 12 (noting that the tool automatically updates AI agent's sources on a daily basis "without any manual work").

III. GOVERNANCE ISSUES IN CONSUMER DIGITAL REPLICATION

The lifecycle of most consumer digital replicas follows a common pattern. Users consent to creation through onboarding and identity verification; the platform uses the provided inputs to train a replica model; the replica is then used—often repeatedly—within the platform's tools and workflows; its outputs are constrained through safeguards and platform rules; access is mediated through pricing and subscriptions; and, if the user later chooses to leave, the replica may be deleted or disabled under conditions set by the platform.[27] At each stage of this lifecycle, control over identity shifts incrementally from the user to the platform. This section will focus on two governance issues that emerge from that shift: the limits of revocable consent over time, and ongoing contests over the permissible scope of use.

A. THE LIMITS OF REVOCABLE CONSENT

Consumer digital replicas often begin with consent practices that reflect the seriousness of what is being authorized.[28] Because replicating a person's voice, likeness, or expressive style implicates core aspects of identity, platforms frequently require more than passive agreement to terms of service. Users may be asked to verify their identity through one-time codes, record themselves reading acknowledgments aloud, or otherwise confirm, in real time, that they are authorizing the use of specific inputs they are uploading. These measures serve an important threshold function: they help ensure that the person whose identity is being replicated is the one giving consent, and that consent is tied to identifiable materials.

That initial authorization, however, rarely stands alone. In addition to enabling the creation of a particular digital replica on the platform, users are often asked whether their inputs may also be used for broader training purposes, such as improving models or refining system performance. At this point, consent begins to operate across time.

27. *See, e.g.*, Jess Diaz-Gomes, *Disagreeable Content*, SYNESTHESIA, https://help.synthesia.io/en/articles/8330530-disagreeable-content [https://web.archive.org/web/20260407025321/https://help.synthesia.io/en/articles/8330530-disagreeable-content] (last visited Mar. 26, 2026) (moderating what type of content an AI avatar can communicate—e.g., legal advice, news reporting, conversations about religious practices—based on subscription plan and degree of specificity); *Customer Terms of Service*, SYNESTHESIA (effective Feb. 23, 2024), https://www.synthesia.io/legal/customer-terms-of-service [https://web.archive.org/web/20260325014043/https://www.synthesia.io/legal/customer-terms-of-service] (reserving discretion not to retain or update custom avatars after termination and committing to deleting them upon account deletion, subject to legal limits).

28. *See, e.g.*, *Create an Avatar*, SYNESTHESIA, https://docs.synthesia.io/docs/personal-avatars [https://web.archive.org/web/20251230064332/https://docs.synthesia.io/docs/personal-avatars] (last visited Mar. 26, 2026) (requiring live consent recording, with the same person in the personal avatar footage creation).

A choice made at onboarding may authorize downstream uses that persist independently of the user's continued engagement with the service.[29]

Crucially, users may later want to disentangle those authorizations. A person might wish to continue using their digital replica—because it is embedded in workflows, accessibility tools, or creative projects—while retracting consent for broader training or reuse beyond that specific replica. Whether such selective withdrawal is possible depends entirely on platform design. Some systems purport to offer granular controls; others do not. In many cases, users are faced with an all-or-nothing choice: retain the replica along with any associated training permissions, or exit the system entirely.[30]

Deletion brings the temporal stakes into sharper relief. When platforms offer deletion, they typically mean deletion of the particular digital replica model associated with the user's account on that platform—disabling access to that replica and, in some cases, removing stored inputs used to create it. What deletion does not always make clear is whether and to what extent training effects derived from those inputs persist elsewhere in the system. A user may be able to stop using a replica without knowing whether their identity data continues to shape other models or features.

This matters because identity-linked training data is not a transient resource. Unlike storage space or computer processing power, trained representations of a person's voice, likeness, or expressive patterns can remain valuable, reusable, and sensitive long after the user's immediate relationship with the platform has ended. Retained training data may be repurposed as systems evolve, exposed in the event of a breach, or applied in contexts the user did not contemplate when consenting.[31] Even where no misuse occurs, continued retention without ongoing authorization raises concerns about autonomy and control that are not resolved simply because a service was once provided in exchange.

The practical significance of these issues is compounded by limits on verification and enforcement. Users must largely trust platform representations about how training works, what deletion reaches, and whether consent has been meaningfully withdrawn. Those claims are difficult to test.[32] Once inputs have been incorporated into training processes—particularly where influence is diffuse or embedded in model parameters—users lack tools to inspect or audit compliance. Even in litigation, challenges remain. Claims about retention or reuse of identity data may be difficult to substantiate without access to internal systems, and obtaining such access typically requires surviving early

29. *See* Jennifer King et al., *User Privacy and Large Language Models: An Analysis of Frontier Developers' Privacy Policies* 1, 6, ARXIV (Sept. 10, 2025), https://arxiv.org/html/2509.05382v1 [https://web.archive.org/web/20260213041359/https://arxiv.org/html/2509.05382v1].

30. *See* Hannah Ruschemeier, *Generative AI and Data Protection* 11, 13, CAMBRIDGE F. ON AI: LAW & GOVERNANCE (2025).

31. *See* King et al., *supra* note 29, at 8–9.

32. *See, e.g.*, Judy Hanwen Shen et al., *The Limits of AI Data Transparency Policy: Three Disclosure Fallacies* 1, 10–12, ARXIV (Jan. 26, 2026), https://arxiv.org/html/2601.18127v1#S5 [https://web.archive.org/web/20260304134349/https://arxiv.org/html/2601.18127v1#S5] (noting that AI data-transparency rules often depend on self-reported disclosures that are difficult to audit or verify in practice—producing an "enforcement gap" that limits meaningful testing of whether companies actually follow their stated data-handling practices).

motions to dismiss.[33] Absent clear evidence of continued retention or use, proving that consent was exceeded can be a significant hurdle. In this sense, revocable consent operates less as a technical guarantee than as a promise whose credibility depends on platform governance and transparency.

B. CONTESTING THE SCOPE OF USE

If consent governs entry into digital replica systems, control over how replicas function day to day is exercised through platform architecture. Safeguards, portability limitations, pricing structures, and evolving contractual terms collectively shape the scope of permissible use.

Safeguards are the most visible expression of this control. They typically operate at the level of prompts and outputs, rather than at the level of training or persistence.[34] The replica continues to exist, but certain expressions or applications are restricted after the fact. In many systems, these safeguards are enforced primarily through automated monitoring. Automation can produce both underinclusive and overinclusive outcomes. In some situations, prohibited uses may slip through; in others, legitimate or advertised uses may be blocked because the system misinterprets intent or context.

Users often respond by adapting. They rephrase prompts, divide tasks into stages, chain multiple tools together, or move parts of a workflow across platforms. These work-arounds may be benign, questionable, or expressly disallowed, but they are a predictable response to systems that promise broad functionality while enforcing it imperfectly. Whether a restriction can be challenged depends on platform infrastructure. Some platforms provide escalation mechanisms—such as human review or expanded usage licenses that permit broader application—while others rely almost entirely on automated enforcement, leaving little room to contest errors or edge cases.

Economic controls further shape the scope of use. Digital replicas are commonly tied to subscription models, usage caps, or tiered access.[35] Tools initially framed as personal conveniences may become relied upon professionally or commercially. As terms of service evolve, continued access to a trained replica may require higher payments, additional licenses, or acceptance of new limitations.[36] In these settings, pricing does more than compensate for services rendered; it regulates who may rely on scalable identity replication and for what purposes.

---

33. *See* Valerian Stolpe, *AI Without Audit Trails Is Becoming a Legal and Governance Liability*, ZL TECH (Dec. 5, 2025), https://www.zlti.com/blog/ai-without-audit-trails-is-becoming-a-legal-and-governance-liability/ [https://web.archive.org/web/20260213071922/https://www.zlti.com/blog/ai-without-audit-trails-is-becoming-a-legal-and-governance-liability/].

34. *See* Markus Anderljung, Julian Hazell & Moritz von Knebel, *Protecting Society from AI Misuse: When Are Restrictions on Capabilities Warranted?,* 40 AI & SOC'Y 3841 (2024).

35. *See* Rupsa Majumdar, Anjula Gurtoo & Minnu Maileckal, *Developing a Data Pricing Framework for Data Exchange*, 11, 4 FUTURE BUS. J. 1, 7 (2025).

36. *See* Nontokozo Mokoena & Ibidun Christiana Obagbuwa, *An Analysis of Artificial Intelligence Automation in Digital Music Streaming Platforms for Improving Consumer Subscription Responses: A Review* 7 FRONTIERS IN AI 2 (2025).

Portability intensifies this contest. At present, most consumer digital replicas are not interoperable. Users can typically export outputs—audio files, images, video clips—and recombine them through traditional editing workflows, but the underlying trained replica usually cannot move across platforms. Identity remains bound to the system that trained it. At the same time, there are clear indications that this may change. Generative AI tools that once required standalone platforms are now embedded directly within creative software ecosystems, allowing users to access similar capabilities across environments.[37] As voice, visual, and behavioral replication converge, comparable expectations of interoperability are likely to emerge for digital replicas themselves.

Until then, users retain alternatives. They can rebuild replicas on different platforms, stitch together outputs using older tools, or dispense with replicas altogether and rely on their real selves—time and stamina permitting. Still, as digital replicas become normalized tools for scalable communication, accessibility, and creation, these alternatives carry increasing cost. The scope of permissible use is therefore not fixed at creation, but continually contested through safeguards, pricing, portability, and design choices that determine how identity may be expressed, constrained, or monetized over time.

IV. CONCLUSION

Digital replicas are no longer a speculative or marginal technology confined to entertainment studios, research labs, or controversial deepfakes. They are increasingly ordinary consumer tools, encountered through products that promise convenience, accessibility, and creative efficiency. The legal challenge they pose is not limited to deception, labor displacement, or misappropriation. It lies in how these systems externalize human identity into persistent, reusable digital forms that operate beyond the moment of human action.

Existing legal definitions of "digital replica" reflect the narrow contexts in which they developed. They are reactive and purpose-built—designed to address preservation, optimization, performer protection, or fraud. Each responds to a real concern, but none captures the broader structural shift now underway. Consumer digital replicas are identity-bearing systems capable of generating new expressive outputs over time. Their legal significance does not depend on realism or intent, but on derivation and continuity: a limited act of contribution can produce an open-ended stream of outputs that remain meaningfully tied to a particular person.

That understanding clarifies why governance concerns arise even where replicas are disclosed, consented to, and self-created. Across voice, visual, and behavioral domains, these technologies are framed as tools rather than as forms of identity delegation.

37. Generative AI video models like Runway, PikaLabs, and OpenAI's Sora are now available directly in video editing software like Adobe Premiere. *See* Jess Weatherbed, *Adobe Premiere Pro Is Getting Generative AI Video Tools,* THE VERGE (Apr. 15, 2024), https://www.theverge.com/2024/4/15/24130804/adobe-premiere-pro-firefly-video-generative-ai-openai-sora [https://web.archive.org/web/20250723005223/https://www.theverge.com/2024/4/15/24130804/adobe-premiere-pro-firefly-video-generative-ai-openai-sora].

Consent is concentrated at onboarding, while platform design governs everything that follows. Over time, control migrates. Decisions about retention, training, safeguards, pricing, and portability are largely set by platforms, not users. Revocability, where it exists, is often partial and difficult to verify.

This gap exposes the limits of existing legal frameworks. Doctrines built around deception, discrete consent, and observable misuse do not map well onto openly synthetic systems that persist and recombine identity over time. The problem is not use—many consumer digital replicas are beneficial—but governance. As these tools become more integrated and normalized, the costs of treating them as ordinary software features will grow. Without a structural account of digital replicas, law arrives only after control over identity has already been decided.

Reframing Deepfakes

Jennifer E. Rothman*

* Jennifer E. Rothman is the Nicholas F. Gallicchio Professor of Law at the University of Pennsylvania Carey Law School. This essay is adapted from a lecture given at Columbia Law School on October 24, 2025, at the Symposium *Deepfakes: In Search of Global Solutions*, co-sponsored by Columbia Law School's Kernochan Center and the Columbia-Sorbonne Alliance. I am thankful for comments from Sarah Boyd, Jane Ginsburg, Michael Goodyear, Jacob Noti-Victor, Elizabeth Pollman, the participants and organizers of the Symposium, and the journal editors, as well as research assistance from Penn Carey Law's library staff, particularly the excellent work by Genevieve Tung, as well as additional research assistance by Rachel Buckland and Shivani Chelliah.

INTRODUCTION

The circulation of deceptive fakes of real people appearing to say and do things that they never did has been made ever easier and more convincing by improved and still improving technology, including (but not limited to) uses of generative artificial intelligence ("AI"). I was asked to speak about the legal landscape in the United States that regulates deepfakes, as well as some predictions of what may be on the horizon in terms of potential future legislation. In some sense, this is a nearly impossible task as hundreds of laws have passed in just the last few years to address concerns over deepfakes. California, for example, seems to be passing new AI-related laws almost every week; six laws were passed in just the last three weeks.[1] So, instead of trying to

1. A.B. 325, 2025 Gen. Assemb., Reg. Sess. (Cal. 2025) (limiting uses of common pricing algorithms); A.B. 489, 2025 Gen. Assemb., Reg. Sess. (Cal. 2025) (regulating undisclosed use of AI tools in healthcare to imply provision of care by "natural person"); S.B. 243, 2025 Gen. Assemb., Reg. Sess. (Cal. 2025) (regulating AI companion chatbots); A.B. 853, 2025 Gen. Assemb., Reg. Sess. (Cal. 2025) (expanding transparency requirements for AI platforms and delaying implementation date of prior transparency provisions); A.B. 621, 2025 Gen. Assemb., Reg. Sess. (Cal. 2025) (expanding prohibition on deepfake pornography); A.B. 316, 2025 Gen. Assemb., Reg. Sess. (Cal. 2025) (preventing civil defendants from avoiding liability by claiming harm was caused by AI tools acting "autonomously"); Kara Williams & Mayu Tobin-Miyaji, *California Tech Legislation Roundup: Numerous Privacy and AI Laws Enacted and Six Vetoed*, EPIC.ORG (Oct. 15, 2025), https://epic.org/california-tech-legislation-roundup-numerous-privacy-and-ai-laws-enacted-and-six-vetoed/ [http://web.archive.org/web/20260206202511/https://epic.org/california-tech-legislation-roundup-numerous-privacy-and-ai-laws-enacted-and-six-vetoed/]. Several of these new bills have particular relevance for deepfakes, including requirements of transparency tools to detect and disclose AI, increasing statutory damages for deepfake pornography, and limiting defenses for AI developers and users.

cover everything, I want to set forth some guideposts for sorting through this increasingly complicated landscape, and only then consider the existing laws that cover identity rights and that likely or explicitly apply to deepfakes, including digital replicas, forgeries, and voice clones.[2] I will also consider some new legislation being contemplated, particularly at the federal level.

As part of these guideposts, I will propose a taxonomy of deepfakes that should guide our understanding of this area of law. Before developing this taxonomy, I will start in Part I by providing a foundational understanding of what we mean by the term "deepfakes," and then in Part II highlight the reasons why we are concerned about them. Both steps are essential before I can construct the taxonomy in Part III.

Developing the proposed taxonomy of deepfakes is a desperately needed task as the urgent calls for legislative fixes to address deepfakes have collapsed distinct types of fakes into one single monolith. This lack of nuance when speaking about deepfakes has sometimes masked the problems at issue and obscured the applicability of existing legal structures to combat them. I divide deepfakes (of humans) into four categories: *unauthorized*; *authorized*; *deceptively-authorized*; and *fictional*. As part of this analysis, I identify the two key considerations for regulating deepfakes, which are (1) whether the fakes are *authorized* by the people depicted in them and (2) whether the fakes *deceive the public* into thinking they are authentic recordings.

In Part IV, I use this taxonomy of deepfakes and our understanding of the harms that potentially flow from each type of deepfake to evaluate whether current and proposed laws address our concerns, exacerbate them, or create new challenges. Unfortunately, much of the recently proposed and enacted legislation does not adequately focus on limiting deceptive deepfakes, and in some instances even legitimizes and incentivizes the creation of them.

I. DEFINING DEEPFAKES

It may seem strange to start by defining the very topic of the symposium. Nevertheless, understanding the scope of what is covered by the term is an essential first step in considering the role of current (and proposed) laws. We must have a sufficiently specific and common definition so we can understand the boundaries of the term and consider whether deepfakes are something substantively new or simply a new manifestation of something longstanding.

The term "deepfakes" is thought to have originated in 2017 with a Reddit user who went by that name and used the term to refer to pornographic images that swapped in the faces of real people to the bodies of others to make it seem like those whose faces were depicted had appeared in the intimate and explicit situations shown when they had not.[3] Since 2017 both the use of the term and its meaning have exponentially

2. I will develop further the meaning of these terms in Part I.

3. Meredith Somers, *Deepfakes, Explained*, MIT SLOAN SCH. MGMT. (July 21, 2020), https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained? [http://web.archive.org/web/20260206203122/https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained] (defining a deepfake as "a specific kind of synthetic media where a person in an image or video is

grown.[4] What was once thought to be primarily a nonconsensual pornography problem has expanded to encompass difficult-to-detect video and voice fakes that simulate or impersonate people in any context. Congressional representatives and some state legislatures have used the language of "digital forgeries" of "identifiable individual[s]" to capture the concept of deepfakes.[5] Some state laws and proposed federal ones, along with the U.S. Copyright Office, have instead adopted the term "digital replicas" and "voice clones" to refer to deepfakes.[6]

The definitions of digital replicas themselves vary. For example, the proposed NO FAKES Act, introduced in April of 2025 in both the U.S. House and Senate, defines a "digital replica" as a "newly-created, computer-generated, highly realistic electronic representation that is readily identifiable as the voice or visual likeness of an

swapped with another person's likeness" and tracing the term's origin to 2017 Reddit user "deepfake"). The Reddit user likely also adopted the term in reference to the "deep" learning technique used by artificial intelligence. *See Deepfake*, ENCYC. BRITANNICA (Feb. 4, 2026), https://www.britannica.com/technology/deepfake [http://web.archive.org/web/20260206203414/https://www.britannica.com/technology/deepfake] (noting that "[t]he term *deepfake* combines *deep*, taken from AI deep-learning technology (a type of machine learning that involves multiple levels of processing), and *fake*, addressing that the content is not real").

4. Although I note that a recent Uniform Law Commission drafting committee narrowed its focus to deepfakes only in the pornographic context after forming a study committee to consider deepfakes more broadly. *Compare* SUZANNE BROWN WALSH & EUGENE VOLOKH, DEEPFAKES STUDY COMM., FINAL STUDY COMMITTEE REPORT 2 (2024), https://www.uniformlaws.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=f0b236e5-f8b6-1825-d0c7-baa88e08256b&forceDialog=0; (considering both election-related deepfakes and those with sexual content), *with* Memorandum from Eugene Volokh to Drafting Comm. on Nonconsensual Pornographic Deepfakes 4 (Dec. 6, 2024) (limiting scope of drafting to uniform law addressing "nonconsensual pornography"), https://www.uniformlaws.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=fd4a4157-1d1c-efce-ceb3-688a6e048b9d&forceDialog=1 .

5. The proposed DEFIANCE Act of 2024 focused on "sexually-explicit 'deepfakes,'" describing them as a "digital forgery," "created through the use of software, machine learning, artificial intelligence, or any other computer-generated or technological means" to falsely appear to be authentic." DEFIANCE Act of 2024, S. 3696, 118th Cong. § 3(a) (2024); *see also* 18 PA. CONS. STAT. § 4101.1 (2025) (creating crime to produce non-consensual "forged digital likenesses," including deepfakes and voice clones, defined as "a computer-generated visual representation of an actual and identifiable individual or audio recording of an actual and identifiable individual's voice"); *see also* Bobby Chesney & Danielle K. Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1758 (2019) (describing "deep fakes" as a form of "realistic and convincing" "digital impersonation").

6. *See, e.g.*, NO FAKES Act of 2025, S.1367, 119th Cong. § 2(a) (introduced Apr. 9, 2025); U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE: PART I: DIGITAL REPLICAS (July 2024); *see also* Preventing Abuse of Digital Replicas Act (PADRA), 118th Cong. (Discussion Draft as introduced by Rep. Darrell Issa, Aug. 9, 2024) (defining a digital replica as a "a computer generated, electronic representation of an identifying characteristic of a subject person, who is an individual human being" but limiting to instances in which the "identifying characteristic" is "distinctive to said subject person" and will be so "associated" with them by reasonable persons in the "relevant industry or market" that the representation is "substantially indistinguishable" and would be "apparent" that "was generated . . . to duplicate" the person (or the identifiable characteristic)).

individual."[7] The Copyright Office defines a digital replica more broadly as "the use of digital technology to realistically replicate an individual's voice or appearance."[8]

The FBI has defined deepfakes as encompassing "a video, an image, or recording convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said."[9] *Merriam-Webster's Dictionary* has adopted a similar definition of a deepfake as "an image or recording that has been convincingly altered and manipulated to misrepresent someone as doing or saying something that was not actually done or said."[10]

Although there are many similarities across these (and other) definitions of deepfakes, there are also significant differences. I want to highlight two crucial ones. First, the definitions differ in whether a deepfake must deceive or be likely to deceive the public that it is an authentic recording or instead simply needs to depict an identifiable person regardless of deception.[11] Some definitions add a requirement that the actor/potential defendant intended to deceive the public by creating or disseminating the deepfake.[12] The requirement of intent to deceive may be driven by concerns that a strict liability or even negligence standard may violate the First Amendment, especially in the context of criminal charges.[13]

A second major difference among the definitions of deepfakes is whether the term only applies to depictions of people or also applies to depictions of objects and places. The European Union's AI Act has adopted a broader vision of deepfakes, as including

---

7. NO FAKES Act of 2025, *supra* note 6, at § 2(a).

8. U.S. COPYRIGHT OFF., *supra* note 6, at "About This Report."

9. *Public Service Announcement: Deepfakes and Stolen PII Utilized to Apply for Remote Work Positions,* FBI INTERNET CRIME COMPLAINT CTR. (IC3) (June 28, 2022), https://www.ic3.gov/PSA/2022/PSA220628 [https://perma.cc/YGA6-PPYY].

10. "*Deepfake*," MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/deepfake [http://web.archive.org/web/20260206231320/https://www.merriam-webster.com/dictionary/deepfake] (last visited Feb. 8, 2026).

11. *Compare* 18 PA. C.S.A. § 4101.1(f)(3) (defining a "forged digital likeness" for purposes of a criminal offense as "a computer-generated visual representation" of a person that "is likely to deceive a reasonable person to believe that the visual representation or audio recording is genuine") *and* WASH. REV. CODE § 42.62.020 (2026) (limiting use of "synthetic media" in political campaigns only if a "reasonable individual" would believe such media to depict a real "appearance, action, or speech"), *with* CAL. CIV. CODE § 1708.86 (2026) (realistic intimate image need not deceive the viewer into thinking depicts real events or authorization by person depicted), *and* TENN. CODE ANN. § 47-25-1103 (Elvis Act of 2024) (tying liability for circulation of digital replicas to their authorization not whether they deceive the public). Notably, New York's postmortem right of publicity statute used to limit liability for uses of digital replicas to those that were "likely to deceive the public" but removed this limitation in late 2025, expanding liability to nondeceptive uses. *See* 2025 N.Y. Laws ch. 616 (S.8391) (signed into law Dec. 11, 2025).

12. *See, e.g.*, TEX. ELEC. CODE § 255.004(e) (2019) (defining deep fake videos as ones that with "intent to deceive . . . appear[] to depict a real person performing an action that did not occur in reality"); TEX. PEN. CODE § 21.165 (2024) (same in context of sexually-explicit content); S.B. 1515, 56th Leg., 2d Reg. Sess. (Ariz. 2024) (proposing election-related regulation of deepfakes that would require intent to harm reputation of person depicted or deceive voters).

13. *Cf.* Counterman v. Colorado, 600 U.S. 66 (2023) (holding that criminal liability for speech requires at least recklessness standard in context of true threats); *Ex parte* Jones, No. PD-0552-18, 2021 WL 2126172 (Tex. Ct. Crim. App. 2021) (reading in narrowing construction of intimate image law in Texas so that requires that person who circulated the image knew or was reckless as to lack of consent).

depictions of "objects, places, entities [and] events" as well as of people.[14] The *Encyclopedia Britannica* also takes this approach and does not limit deepfakes to those that depict human beings.[15]

For purposes of this Lecture, I am going to focus on deepfakes of people rather than those of objects or places both because this has been the focus of concern and legislation in the U.S. and also because we have hundreds of years of various laws that focus on unauthorized uses of a person's identity, particularly their names and likenesses, that apply or may apply to such uses in the context of deepfakes.

I also adopt a definition of deepfakes that does not require demonstration that a viewer or listener has been deceived. The deepfake must appear to be an authentic recording of the person depicted but need not do so deceptively. Nor does appearing authentic require that the alleged recording seem realistic or be in a realistic context.

With these central determinations, my operative definition of a deepfake is: "An image, sound, or performance that depicts a person and appears to be an authentic recording of that person when it is not one." Note that I use "depict" here to encompass both use of a person's likeness and their voice.[16] In spite of the etymological origins of the verb "to depict" to indicate visual depictions, there is no alternative word that captures the meaning as well for audio depictions so I use the term to mean both. In addition, note that this definition includes deepfakes made using any technology, whether computer-generated or not, and without regard to whether AI is used. There is variation across definitions on this point too, but I think the best approach is to be technologically neutral.

I am not suggesting that we adopt this definition formally or write it into legislation. Instead, I view the definition as a starting point that we may wish to further refine. I also am not committed to the terminology "deepfake" but it is a term frequently employed and is the chosen term for this symposium. And if a deepfake is understood as a more neutral term than it sometimes is, it has some advantages conceptually over "digital replicas," "voice clones," or "digital forgeries." Deepfakes can be understood as encompassing a wider variety of uses than these targeted, alternative terms, including uses that are not digital, and that are not rooted to a particular context, like the entertainment industry, or a particular legal frame, like criminal conduct.

Having a working definition of deepfakes is essential for understanding and teasing apart the different types of deepfakes that so often have been either conflated or insufficiently distinguished. Before undertaking that task, however, I want to briefly sketch out some of the harms that flow from deepfakes that depict people.

---

14. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, Laying Down Harmonised Rules on Artificial Intelligence, art. 3(60) 2024 O.J. (L 1689) 1 [EU AI Act]. The Uniform Law Commission (ULC)'s recent Study Committee on Deepfakes initially also took such a broad approach defining "'deepfakes' as shorthand . . . to refer to all video, photographic, and audio forgeries." FINAL STUDY COMMITTEE REPORT, *supra* note 4, at 2.

15. *See Deepfake*, *supra* note 3.

16. Likeness, if interpreted broadly to be a representation of a person's identity, could include voice, but it is useful to separate them for clarity as likeness is most often understood as pointing to a person's visual features.

II. HARMS OF DEEPFAKES

I suspect that the harms created by deepfakes are familiar to this audience. They are nevertheless important to remind ourselves of because, particularly in the United States, sometimes these concerns get lost when drafting legislation with a focus on generating or protecting income for particular industries, such as the recording industry, internet platforms, or technology companies. Failure to know what problems we seek to address makes it impossible to evaluate the sufficiency of the already existing legal frameworks.

In addition, although deepfakes is a term that is most often used pejoratively, not all deepfakes should be prohibited or subjects of legal liability. Distinguishing between when the law should step in and when it should step back in the context of deepfakes requires a more nuanced understanding of what harms flow from the dissemination of some (but not all) deepfakes.

There are three main categories of harms that flow from the dissemination of deepfakes: (a) ones that negatively impact the individuals depicted; (b) those that cause broader harms to the public, particularly by deceiving the public into thinking the fakes are real; and (c) those that negatively affect either those close to the person depicted or those with financial interests entangled with the person depicted. Not all of these harms deserve legal redress but those that do largely arise in two main contexts: (1) when they are *not authorized* by the person depicted; or (2) when they *deceive the public* as to their authenticity.

Sometimes deepfakes are both unauthorized and deceptive and at other times they may be one without the other. In each of these instances, some harms are likely to flow. I will largely focus on the harm to living people but will note when relevant issues are raised by deepfakes that portray the deceased.[17] If a deepfake is authorized and not deceptive, it is presumptively benign; even though such fakes may lead to market disruption, these harms are generally not appropriate for legal remediation.

A. HARMS TO PERSON DEPICTED

Unauthorized deepfakes, like other unauthorized uses of a person's identity, can cause a host of personal and economic injuries to the person depicted.[18] Losing control over one's own name, image, likeness, and voice harms our right of self-determination and autonomy. There is a longstanding understanding that we each have both liberty-based and property-based rights in our own names, likenesses, and identity more

17. I note that I use the term "portray" to include both audio and visual depictions, even though the word "portray," like "depict," stems etymologically from a visual context.

18. JENNIFER E. ROTHMAN, THE RIGHT OF PUBLICITY: PRIVACY REIMAGINED FOR A PUBLIC WORLD 98–112 (2018) (considering and evaluating various justifications for protections against unauthorized uses of a person's identity); Robert C. Post & Jennifer E. Rothman, *The First Amendment and the Right(s) of Publicity*, 130 YALE L.J. 86, 93–125 (2020).

broadly.[19] Unauthorized deepfakes may also cause humiliation and degradation that injures a person's dignity and reputation. This is particularly so when a person is placed in troubling contexts, such as pornography or saying problematic or offensive things that they never said.[20] This harm exists even if the use does not deceive the public as to its authenticity, though the scale of the harms may differ.[21]

The person depicted may also suffer a range of market-based harms from "lost job opportunities and endorsement deals to reduced salaries, loss of revenue from licensing and merchandising contracts, and overall diminishment of [their] goodwill."[22] This is particularly true (but not exclusively so) for those who actively commercialize their identities. Unauthorized deepfakes, particularly of performances, could substitute for hiring the person themselves and, if not regulated, could reduce incentives for actors, singers, and other performers to create such performances in the first place.[23]

In the context of deceased individuals who are depicted in deepfakes, they can no longer suffer direct personal injuries. Nevertheless, we might consider injuries to the living (what I elsewhere refer to as "future-decedents") who may be troubled by their own potential postmortem treatment in deepfakes. The consideration of postmortem identity rights and deepfakes that depict the deceased is a large topic unto itself and one that I cannot do justice to here, but one that I have considered in depth elsewhere.[24]

---

19. *See* Vidal v. Elster, 602 U.S. 286 (2024) (considering long history of protecting a person's name under trademark and unfair competition laws); Pavesich v. New Eng. Life Ins. Co., 50 S.E. 68, 76–77 (Ga. 1905) (concluding that rights over a person's likeness is both a liberty and property right); Roberson v. Rochester Folding Box Co., 64 N.E. 442, 449–50 (N.Y. 1902) (Gray, J., dissenting) (suggesting that every person has a "property right" in their own "person" and image); Corliss v. E.W. Walker Co., 64 F. 280, 282 (C.C.D. Mass. 1894) (recognizing that every person has a property right in their portrait); ROTHMAN, *supra* note 18, at 11–44; Jennifer E. Rothman, *Navigating the Identity Thicket: Trademark's Lost Theory of Personality, the Right of Publicity, and Preemption*, 135 HARV. L. REV. 1271, 1297 (2022); *The Right to Privacy*, 6 GREEN BAG 498, 499 (1894) (suggesting that there is "a right of property in one's personal appearance"); *see also* Anita L. Allen & Jennifer E. Rothman, *Postmortem Privacy*, 123 MICH. L. REV. 285, 297–98, n.53 (2024) (considering framing of publicity and privacy rights arising out of a person's identity as a property right in context of postmortem privacy).

20. *See* Post & Rothman, *supra* note 18, at 121–25, 165–71; *see also* Chesney & Citron, *supra* note 5, at 1771–75 (pointing to harms from deepfakes arising from extortion, sexual "exploitation," and "sabotage"); Michael P. Goodyear, *Dignity and Deepfakes*, 57 ARIZ. ST. L.J. 931, 942–53 (2026) (focusing on dignitary injuries caused by being portrayed in deepfakes, particularly though not exclusively in the context of sexualized depictions). Some states have extended postmortem rights targeting depictions of soldiers specifically to preserve the dignity of the deceased and their surviving relatives. *See, e.g.*, ARIZ. REV. STAT. § 12-761; ARIZ. REV. STAT. § 13-3726; LA. REV. STAT. ANN. § 14:102.21; *see also* Allen & Rothman, *supra* note 19, at 318–20, 322–24 (considering dignity-based reasons for protecting postmortem rights).

21. *See* discussion *infra* Part III.A.

22. Post & Rothman, *supra* note 18, at 108 (citing WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 222–28 (2003); ROTHMAN, *supra* note 18, at 110–11; Mark F. Grady, *A Positive Economic Theory of the Right of Publicity*, 1 UCLA ENT. L. REV. 97, 103–04 (1994)).

23. This is particularly relevant when the underlying performance is not protected by copyright. *See, e.g.*, Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562 (1977). In such instances, the objectives of identity rights and copyright law overlap. *See id.* at 575; Post & Rothman, *supra* note 18, at 96–106.

24. *See* Allen & Rothman, *supra* note 19, *passim* (considering a variety of interests that could support postmortem publicity rights, including the interests of "future-decedents," the "relational-living," and of society more generally); *see also* ROTHMAN, *supra* note 18, at 81–86 (considering expansion of right of publicity laws to cover postmortem rights); Jennifer E. Rothman, *Postmortem Publicity Rights at the Property-Personality*

B. HARMS TO THE PUBLIC

Deceptive deepfakes—whether authorized or not by the person depicted—also can cause significant harm to the public. They can destabilize our political system by circulating fake images and recordings of political figures saying and doing things they never did in ways that can affect how voters perceive them and alter the outcome of elections. Deepfakes of politicians could cause civil unrest and even global catastrophes by inciting wars or conflicts engendered by false statements or actions appearing to be authentic speech or conduct by a world leader.[25]

Deceptive deepfakes can also more broadly destabilize our access to information and truth. As Brian Chen recently wrote in the *New York Times*, we may be facing the "end of visual fact."[26] Can society survive if we not only do not have common references and sources, but also do not have reliable documentation of real-world events? The criminal justice system and the tort system will themselves be threatened by the undermining of image-and-voice-based evidence.[27]

We as a society may also be impoverished by AI-generated slop in the place of high-quality content. This could happen with nondeceptive deepfakes too, but as long as the public is able to knowingly choose between deepfakes and authentic performances this is not something the law should generally attend to when the uses are otherwise authorized.[28] But it is reasonable to construct a legal regime that supports the public knowingly choosing to watch AI-generated performances rather than consuming them thinking they are watching or listening to real actors and singers.[29]

Deceptive deepfakes can also disrupt consumer markets, leading people to make purchasing decisions based on false information about what a depicted person endorses or uses.

Divide, *in* PRIVATE LAW THEORY & INTELLECTUAL PROPERTY (Shyamkrishna Balganesh, Poorna Mysoor & Henry Smith eds., Cambridge Univ. Press forthcoming 2027), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4971180 (suggesting that if we conceptualize the right of publicity as a property right, then its unique status indicates that any postmortem right is best understood as something new rather than as a descendible right).

25. For a useful analysis of some of these harms, *see* Chesney & Citron, *supra* note 5, at 1771–86.

26. Brian X. Chen, *A.I. Video Generators Are Now So Good You Can No Longer Trust Your Eyes*, N.Y. TIMES, Oct. 9, 2025, https://www.nytimes.com/2025/10/09/technology/personaltech/sora-ai-video-impact.html [https://web.archive.org/web/20251009113425/https://www.nytimes.com/2025/10/09/technology/personaltech/sora-ai-video-impact.html].

27. *See* Rebecca A. Delfino, *Deepfakes on Trial: A Call to Expand the Trial Judge's Gatekeeping Role to Protect Legal Proceedings from Technological Fakery*, 74 HASTINGS L.J. 293, 297 (2023).

28. Such authorization may need to include authorization by copyright holders, as well as by the person(s) depicted. AI outputs may infringe copyrighted works and it is an open question of whether training AI models using copyrighted materials and people's identities is infringing of copyrights and publicity rights. *See infra* note 70 and accompanying text.

29. *See* Jacob Noti-Victor, *Regulating Hidden AI Authorship*, 111 VA. L. REV. 139 (2025).

C. HARMS TO RELATED PARTIES

Those close to those depicted may also suffer emotional distress, such as a parent seeing their child depicted in deepfake pornography, or a relative seeing their deceased loved one reanimated against their wishes. Unauthorized deepfakes could also disrupt the income for companies that have invested in a person's performance or that hold the copyrights to various works in which the person appears that may have been used to create the deepfakes. This is particularly a risk when the fakes might substitute for paying for authentic works by the underlying person.

* * *

This overview of the harms that flow from deepfakes highlights that the key determinants of whether deepfakes cause harm stem from whether the deepfakes are *unauthorized* or *deceptive* or both. This focus informs the development of the taxonomy that follows.

III. A TAXONOMY OF DEEPFAKES

To the extent deepfakes are distinguished from one another it has primarily been on the basis of the context in which the fakes appear—for example, to distinguish among deepfakes that appear in the context of political campaigns or that depict politicians, those that show private body parts or are otherwise pornographic, and those that impersonate well-known performers.

These contextual distinctions have obscured deeper thinking about whether the deepfakes across these and other contexts are (or should be) different from one another from a jurisprudential perspective. A more nuanced parsing of deepfakes is essential to better distinguish between the problems that are appropriate for legal redress versus those that are more appropriate for collective bargaining or market-based solutions, or that simply must be tolerated and in some instances even celebrated. The focus on the context in which deepfakes appear rather than on their other characteristics has also exacerbated the spread of overlapping and sometimes conflicting laws that cover a person's identity.[30] These contextually-targeted laws also may be less likely to survive constitutional challenges in part because their contextual targeting may make them underinclusive at addressing the problems at hand.[31]

I divide deepfakes (of humans) into four categories: *unauthorized*; *authorized*; *deceptively-authorized*; and *fictional*.

30. *See* Rothman, *supra* note 19, at 1278–88 (describing some of these overlapping laws before the recent explosion in AI-related legislation that has worsened the problem).

31. *See, e.g.*, Kohls v. Bonta, 797 F.Supp.3d 1177 (E.D. Cal. 2025) (striking down as unconstitutional recent California law regulating deepfakes in the election context in part because was underinclusive and viewpoint discriminatory); *cf.* 281 Care Committee v. Arneson, 766 F.3d 774 (8th Cir. 2014) (holding unconstitutional a Minnesota statute that made it a crime to knowingly or with reckless disregard for the truth make a false statement about a proposed ballot initiative in part because was underinclusive).

A. UNAUTHORIZED DEEPFAKES

Most of the expressed concerns over deepfakes have centered on the unauthorized use of a person's likeness, voice, or performance in ways that they never agreed to and that could deceive viewers or listeners into thinking that the person actually appeared or performed in the disseminated works. Recent calls for legislative action around deepfakes have stemmed from high-profile examples of such unauthorized uses, including the 2023 viral AI-generated song, "Heart on My Sleeve," which imitated the voices of successful recording artists Drake and The Weeknd.[32] The song became a hit and people thought it was an authentic new release from the famous artists.[33] Numerous other recording artists and actors have found themselves depicted in deepfakes. They are often shown falsely endorsing products, such as Tom Hanks being deceptively used in an ad for dental services, Taylor Swift "peddl[ing]" Le Creuset Cookware, or the recent fake Will Ferrell Doritos ad.[34]

The famous and the ordinary have been depicted by classmates, former partners, and strangers in pornographic contexts that they never appeared in using generative AI technology.[35] Politicians too have been victims of unauthorized and deceptive uses

32. *See* THE DAILY, *The Ballad of "Deepfake Drake"* (N.Y. Times Podcasts, Apr. 28, 2023), https://www.nytimes.com/2023/04/28/podcasts/the-daily/ai-deepfake-drake.html [https://web.archive.org/web/20260207024948/https://www.nytimes.com/2023/04/28/podcasts/the-daily/ai-deepfake-drake.html]; Chris Willman, *AI-Generated Fake "Drake"/"Weeknd" Collaboration, "Heart on My Sleeve," Delights Fans and Sets Off Industry Alarm Bells*, VARIETY (Apr. 17, 2023), https://variety.com/2023/music/news/fake-ai-generated-drake-weeknd-collaboration- heart-on-my-sleeve-1235585451/ [https://web.archive.org/web/20260207025005/https://variety.com/2023/music/news/fake-ai-generated-drake-weeknd-collaboration-heart-on-my-sleeve-1235585451/].

33. *See* Willman, *supra* note 32.

34. Michaela Zee, *Tom Hanks Warns Fans About "AI Version of Me" Promoting Dental Plan: "I Have Nothing to Do With It,"* VARIETY (Oct. 1, 2023), https://variety.com/2023/film/news/tom-hanks-ai-video-dental-plan-warns-fans-1235741781/ [https://web.archive.org/web/20260207025138/https://variety.com/2023/film/news/tom-hanks-ai-video-dental-plan-warns-fans-1235741781/]; Tiffany Hsu & Yiwen Lu, *No, That's Not Taylor Swift Peddling Le Creuset Cookware*, N.Y. TIMES (Jan. 9, 2024), https://www.nytimes.com/2024/01/09/technology/taylor-swift-le-creuset-ai-deepfake.html [https://web.archive.org/web/20260220181022/https://www.nytimes.com/2024/01/09/technology/taylor-swift-le-creuset-ai-deepfake.html]; *see also* ASTRO DYNAMICS, *Will Ferrell Doritos Super Bowl 2025 Ad* (YouTube, Feb. 2, 2025), https://www.youtube.com/watch?v=CLCy3H1ABj0 [https://web.archive.org/web/20260207025118/https://www.youtube.com/watch?v=CLCy3H1ABj0] (AI-generated fake ad not authorized by Ferrell or Doritos).

35. *See, e.g.*, Ashley Belanger, *NJ Teen Wins Fight to Put Nudify App Users in Prison, Impose Fines up to $30k*, ARS TECHINICA (Apr. 4, 2025), https://arstechnica.com/tech-policy/2025/04/adults-told-her-to-move-on-instead-teen-won-fight-to-criminalize-deepfakes/ [https://web.archive.org/web/20260207025623/https://arstechnica.com/tech-policy/2025/04/adults-told-her-to-move-on-instead-teen-won-fight-to-criminalize-deepfakes/] (describing such an incident involving teens in New Jersey, as well as legislation passed to address it within the state); Imran Rahman-Jones, *Taylor Swift Deepfakes Spark Calls in Congress for New Legislation*, BBC (Jan. 27, 2024), https://www.bbc.com/news/technology-68110476 [https://web.archive.org/web/20260207025704/https://www.bbc.com/news/technology-68110476]; Rachel DeSantis, *Kristen Bell Recalls Shock of Learning her Face Was Used in Pornographic Deepfake: "It's Not OK,"*

of their identities, including Presidents Joe Biden and Barack Obama, Senator Amy Klobuchar, and U.K. Prime Minister Rishi Sunak.[36] Deepfakes, particularly voice clones, have been used to deceive family members into thinking their relative is in danger, leading them to pay out large sums to criminals.[37] This technology is getting better every day, and Open AI's recent release of Sora 2 has generated significant concerns with its swift ability to generate authentic-seeming performances using the identities of real actors (and anyone else) based on ingesting existing performances.[38]

PEOPLE (June 10, 2020), https://people.com/human-interest/kristen-bell-shock-face-used-pornographic-deepfake/ [https://web.archive.org/web/20260207025757/https://people.com/human-interest/kristen-bell-shock-face-used-pornographic-deepfake/].

36. *See* Cristina Criddle, *Political Deepfakes Top List of Malicious AI Use, DeepMind Finds*, FIN. TIMES (June 25, 2024), https://www.ft.com/content/8d5bc867-c69d-44df-839f-d43c92785435 [https://web.archive.org/web/20260207025801/https://www.ft.com/content/8d5bc867-c69d-44df-839f-d43c92785435]; Joan Donovan & Britt Paris, *Beware the Cheapfakes*, SLATE (June 12, 2019), https://slate.com/technology/2019/06/drunk-pelosi-deepfakes-cheapfakes-artificial-intelligence-disinformation.html [https://web.archive.org/web/20260207025807/https://slate.com/technology/2019/06/drunk-pelosi-deepfakes-cheapfakes-artificial-intelligence-disinformation.html] (describing impact of low tech fakes, including "Drunk Pelosi," in which a video was slowed down to produce the effect of making it seem that Representative Pelosi was drunk when she was not); Victoria Elliott & Makena Kelly, *The Biden Deepfake Robocall Is Only the Beginning*, WIRED (Jan. 23, 2024), https://www.wired.com/story/biden-robocall-deepfake-danger/ [https://web.archive.org/web/20260207025953/https://www.wired.com/story/biden-robocall-deepfake-danger/] (describing a voice clone of President Biden that discouraged voting); Amy Klobuchar, *What I Didn't Say About Sydney Sweeney*, N.Y. TIMES (Aug. 20, 2025), https://www.nytimes.com/2025/08/20/opinion/amy-klobuchar-deepfakes.html [https://web.archive.org/web/20260207025913/https://www.nytimes.com/2025/08/20/opinion/amy-klobuchar-deepfakes.html]. This concern over the political dangers of deepfakes is not new to this generative AI moment. It was well illustrated by Jordan Peele's use of the technology to fake authentic videos of former President Barack Obama, intended to warn of the potential for inciting a war if world leaders are portrayed saying things they never said that could potentially launch wars. *See* Hallie Jackson, *Fake Obama Warning About "Deep Fakes" Goes Viral*, MS NOW (Apr. 19, 2018), https://www.ms.now/hallie-jackson/watch/fake-obama-warning-about-deep-fakes-goes-viral-1214598723984 [https://web.archive.org/web/20260207030047/https://www.ms.now/hallie-jackson/watch/fake-obama-warning-about-deep-fakes-goes-viral-1214598723984].

37. *See, e.g.*, Charles Bethea, *The Terrifying A.I. Scam That Uses Your Loved One's Voice*, NEW YORKER (Mar. 7, 2024), https://www.newyorker.com/science/annals-of-artificial-intelligence/the-terrifying-ai-scam-that-uses-your-loved-ones-voice [https://web.archive.org/web/20260301044605/https://www.newyorker.com/science/annals-of-artificial-intelligence/the-terrifying-ai-scam-that-uses-your-loved-ones-voice].

38. THE TOWN WITH MATTHEW BELLONI, *Sora 2, AI Actors, and How Hollywood Can Fight Back* (Puck, Oct. 2, 2025), https://puck.news/podcast_episode/sora-2-ai-actors-and-how-hollywood-can-fight-back/ [https://web.archive.org/web/20260207030127/https://puck.news/podcast_episode/sora-2-ai-actors-and-how-hollywood-can-fight-back/]. Since the time of the lecture, as predicted, many other models have sprung up, notably Seedance 2.0, which recently received significant attention for its use to generate a short of movie stars Tom Cruise and Brad Pitt fighting and talking about Jeffrey Epstein. *See, e.g.*, Jake Kanter, *Cruise vs Pitt Deepfake: TikTok Owner's New AI Video Model Appears to Be Regurgitating Hollywood Movies on Epic Scale*, DEADLINE (Feb. 12, 2026), https://deadline.com/2026/02/cruise-vs-pitt-seedance-viral-ai-hollywood-videos-1236717127/ [https://web.archive.org/web/20260315013551/https://deadline.com/2026/02/cruise-vs-pitt-seedance-viral-ai-hollywood-videos-1236717127/]; FILMY ATTACK, *AI Brad Pitt vs Tom Cruise Brawl Shocks Internet!* (YouTube, Feb. 12, 2026), https://www.youtube.com/watch?v=noz_aofEpcs [https://perma.cc/ZHY6-S2ZH].

These unauthorized deepfakes can cause all of the harms identified in Part II, including personality-based and market-based harms to the person depicted. Some deepfakes will make clear that they are fake through disclosures, absurd situations, or unrealistic depictions or other tells. But even in these contexts the public could be deceived into thinking the fakes are authorized or sponsored by the person depicted. Regardless of whether the public is deceived, the person depicted may be exposed to the same personal and financial harms. Those who are financially or otherwise connected to the person depicted may also suffer harms from such unauthorized deepfakes without regard to whether they deceive the public.

It is important to note, however, that not all unauthorized deepfakes should be barred. This is particularly so when the public is not deceived as to their authenticity and the uses further significant speech values such as telling creative stories that reference historical events or that provide political commentary. Consideration of possible First Amendment and other speech-related defenses to legal claims arising out of uses of deepfakes is beyond the scope of this Lecture, but is important to flag.[39] It is also worth highlighting that such speech protections will be at their nadir if the deepfakes are likely to deceive the public into thinking they are authentic because rather than serving the objectives of the First Amendment in such instances, the deepfakes undermine public discourse and the search for truth by destabilizing the public's perception of actual events. First Amendment protections are also likely to be limited when the uses could substitute for the work of the person depicted, particularly in the context of performances.[40]

B. AUTHORIZED DEEPFAKES

Despite their dominance in the public narrative, not all deepfakes are unauthorized. In some instances, the people depicted have authorized the use of their voices, likenesses, or performances. These authorized deepfakes therefore do not pose injuries to the person depicted but they may still pose harms to the public if they are deceptive. When they do not deceive the public, these authorized deepfakes do not cause cognizable harms and should be allowed.

Authorized deepfakes are becoming increasingly common. The company Metaphysic, for example, creates digital replicas of real people with the permission and coordination of the underlying person. Metaphysic created a replica of the rap star

39. For an in-depth consideration of the First Amendment analysis in the context of right of publicity claims, *see* Post & Rothman, *supra* note 18 *passim*; *see also* ROTHMAN, *supra* note 18, at 138–59.

40. *See* Post & Rothman, *supra* note 18, at 102–05, 146–48; Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562, 575–76 (1977); *In re* NCAA Student-Athlete Name & Likeness Licensing Litig., 724 F.3d 1268 (9th Cir. 2013); Hart v. Elec. Arts, Inc., 717 F.3d 141 (3d Cir. 2013); *see also* Oral Argument at 9:26, *In re NCAA*, 724 F.3d 1268 (No. 10-15387), https://www.ca9.uscourts.gov/media/view_video.php?pk_vid=0000006196 [https://web.archive.org/web/20260207030144/https://www.ca9.uscourts.gov/media/video/?20120713/10-15387/] (Judge Bybee expressing concern that if avatars of athletes were allowed in a video game then a movie studio might be able to reanimate Tom Cruise in a new *Mission Impossible* movie without hiring him to perform).

Eminem for use in his music video for his song *Houdini*.[41] The video shows Eminem dancing with a younger, digital-replica version of himself.[42] Metaphysic also worked with film director Robert Zemeckis to create de-aged versions of the actors Tom Hanks and Robin Wright for the 2024 movie *Here*.[43] YouTube has been working on voice clone technology that allows users to generate new songs that are voiced by famous singers. The company has been testing this technology as part of its Dream Track tool and has obtained permission from Charlie Puth and at least eight other artists including John Legend, Sia, and Charlie XCX to use their voices in these new AI-generated songs. The tracks sound as if these artists created new performances even though the sound files are entirely digitally generated.[44] Outside of entertainment, authorized uses are also growing. Speechify is a text-to-speech reader initially developed to help those with dyslexia and other learning differences to access written works more easily. The company has partnerships with a number of celebrities, including Snoop Dogg and Gwyneth Paltrow, and uses voice clones of them to make it seem as if these well-known performers are reading texts aloud to the listeners.[45]

Each of these uses is authorized and each of these examples have been used in ways that are either disclosed to the audience that receives them or are used in ways that are unlikely to deceive those who see or hear them. This need not be the case, however. Authorized deepfakes could deceive their audience. When they do, the law should address them because these fakes could cause the same set of harms to the public as unauthorized uses. On the other hand, we should leave room for creative and beneficial uses of nondeceptive deepfakes when authorized by the people depicted.

41. Damien Scott, *How Eminem Used AI to Bring Slim Shady Back to Life*, BILLBOARD (July 17, 2024), https://www.billboard.com/music/rb-hip-hop/eminem-slim-shady-houdini-video-making-of-1235732708/ [https://web.archive.org/web/20260217072438/https://www.billboard.com/music/rb-hip-hop/eminem-slim-shady-houdini-video-making-of-1235732708/].

42. EMINEMMUSIC, *Eminem—Houdini [Official Music Video]*, (YouTube, May 31, 2024), https://www.youtube.com/watch?v=22tVWwmTie8 [https://web.archive.org/web/20260301151713/https://www.youtube.com/watch?v=22tVWwmTie8].

43. Benj Edwards, *New Zemeckis Film Used AI to De-Age Tom Hanks and Robin Wright*, ARS TECHNICA (Nov. 24, 2024), https://arstechnica.com/ai/2024/11/new-zemeckis-film-used-ai-to-de-age-tom-hanks-and-robin-wright/ [https://web.archive.org/web/20260221220936/https://arstechnica.com/ai/2024/11/new-zemeckis-film-used-ai-to-de-age-tom-hanks-and-robin-wright/].

44. Eileen AJ Connelly, *YouTube Reveals "Dream Track," an Al Music Generator Using Nine Famous Singers' Sounds*, WRAP (Nov. 16. 2023), https://www.thewrap.com/youtube- dream-track-explained-ai-songs-charlie-puth-demi-lovato/ [https://web.archive.org/web/20260207030200/https://www.thewrap.com/youtube-%20dream-track-explained-ai-songs-charlie-puth-demi-lovato/]; YOUTUBE, *Introducing Dream Track—an experiment on YouTube Shorts—featuring Charlie Puth*, (YouTube, Nov. 14, 2023), https://www.youtube.com/watch?v=1gjuHUy0IMM [https://web.archive.org/web/20260207030307/https://www.youtube.com/watch?v=1gjuHUy0IMM].

45. SPEECHIFY, https://speechify.com/ [https://web.archive.org/web/2/https://speechify.com/] (last visited Feb. 20, 2026).

C. DECEPTIVELY-AUTHORIZED DEEPFAKES

A third category of deepfakes is often overlooked but essential to understand as a distinct category. Here a person may have agreed to appear in one work or recording but they did not agree to have their voice, likeness, or performance reused in a new context, such as a deepfake. Or, alternatively, the depicted person may not own or control the rights to their own name, likeness, or voice. In each of these scenarios, deepfakes might be categorized as "authorized" in a legal sense but in fact are "unauthorized" in the most important sense because the person whose voice or image is used in the deepfake did not knowingly approve of the specific use. I designate these deepfakes as *deceptively-authorized deepfakes*. These deceptively-authorized deepfakes cause the very same harms to the depicted person that form the basis for regulating deepfakes in the first place. These deceptively-authorized deepfakes also presumptively deceive the public into thinking the person depicted authorized them.

In prior work, I have questioned the legitimacy and constitutionality of allowing someone other than the person themselves—what I have dubbed the "identity-holder"—to own that person's name, likeness, or voice.[46] I have also warned about allowing broad licenses that would give long-term and extensive control over a person's identity to someone other than the identity-holder.[47] Yet, some new and longstanding state laws suggest such transfers and broad licenses may be possible, and several recently proposed bills in Congress to address deepfakes allow someone other than the identity-holder to own or control that person's digital replica.[48] Minors, student-athletes, aspiring actors, recording artists, and models may be particularly vulnerable to having others take control, or even ownership of their voices, likenesses, and performances.[49] But each of us is also at risk as we agree to online terms of service that we do not read

46. ROTHMAN, *supra* note 18, at 115–37; Jennifer E. Rothman, *The Inalienable Right of Publicity*, 101 GEO. L.J. 185, 191 (2012); *see also* Rothman, *supra* note 19, at 1309–17, 1325–31 (considering challenges to the transferability of personal marks in the context of trademark law).

47. *See* Rothman, *supra* note 46, at 234–36; Jennifer E. Rothman, *Reintroduced No FAKES Act Still Needs Revision*, REGUL. REV. (Aug. 18, 2025) (critiquing proposed licensing regime which does not adequately protect identity-holders), https://www.theregreview.org/2025/08/18/rothman-reintroduced-no-fakes-act-still-needs-revision/ [https://web.archive.org/web/20260207030338/https://www.theregreview.org/2025/08/18/rothman-reintroduced-no-fakes-act-still-needs-revision/].

48. *See, e.g.*, TENN. CODE ANN. § 47-25-1103 (stating that property rights to a person's "name, photograph, voice, or likeness" are all "freely assignable"); NO FAKES Act of 2025, *supra* note 6 (proposing ten-year licensing terms with insufficient limits); No AI FRAUD Act, H.R. 6943, 118th Cong. (2024) (authorizing wholesale transfer of a person's rights to their own voices and likenesses in context of digital replicas).

49. *See Artificial Intelligence and Intellectual Property: Part II—Identity in the Age of AI: Hearing Before the Subcomm. on Cts., Intell. Prop. & the Internet*, 118th Cong. 9–10 (2024) (Statement of Jennifer E. Rothman), available at https://rightofpublicityroadmap.com/wp-content/uploads/2024/02/Rothman_Statement_Subcommittee-on-IP_February-2_2024_Submitted.pdf [https://web.archive.org/web/20260207030456/https://rightofpublicityroadmap.com/wp-content/uploads/2024/02/Rothman_Statement_Subcommittee-on-IP_February-2_2024_Submitted.pdf].

and that claim to be able to use our images and recordings in new contexts, including deepfakes, without our permission.[50]

Deceptively-authorized deepfakes raise complicated questions about the intersection of a variety of legal regimes, including contract law, state publicity rights, and federal copyright law. One recent case, *Lehrman v. Lovo*, raises this issue.[51] The company Lovo reached out to two voiceover actors and paid them to record and then use their voices in the development of its text-to-speech software. The actors were told the uses would be for "research purposes only."[52] The company, however, went on to clone the actors' voices and use them more broadly without obtaining additional permission or paying them for the uses—even asserting to customers that the uses were authorized and the outputs were legitimately available for use by their customers.[53] This case could be understood as an unauthorized deepfake case, but the initial voice clones were created with authorization and to the extent such initial captures were protected by copyright, copyright could enable such uses.[54] We will increasingly see cases like this in which a person agreed to one use but not another.[55]

Even though the scale and nature of the problem is growing because a person's performance can so easily be replicated and in ways that are difficult to detect, this problem is longstanding.[56] If a person gives permission to be captured in a copyrighted work, then this work can be reused, including in derivative works, under federal law. This circumstance has led to a number of lawsuits, raising the question of whether state right of publicity laws—that prohibit unauthorized uses of a person's name, voice, or likeness—are preempted by copyright law in such instances. Consider the *Laws v. Sony Music* case, in which a singer's recording was reused in a new recording without her additional permission. The Ninth Circuit Court of Appeals concluded that copyright

50. *See, e.g.*, *Terms of Service*, X, https://x.com/en/tos [https://perma.cc/7ZWZ-J5YE] (last visited Feb. 20, 2026) (providing license to X to use, adapt, modify, and transform any media uploaded to platform); *Terms of Use*, INSTAGRAM, https://help.instagram.com/termsofuse [https://perma.cc/MM8Y-9REP?type=image] (last visited Mar. 1, 2026) (providing license to modify media uploaded to platform, in addition to the right to "prepare derivative works"); *see also Artificial Intelligence and Intellectual Property*, *supra* note 49, at 9–10.

51. Lehrman v. Lovo, Inc., 790 F. Supp. 3d 348 (S.D.N.Y. 2025).

52. *Id.* at 356.

53. *Id.*

54. *Id.* at 381–82, 388 (allowing right of publicity/privacy claims to proceed beyond the motion to dismiss stage under N.Y. Civil Rights Law §§ 50–51).

55. *Cf.* Complaint, Vacker v. Eleven Labs, No. 1:24-cv-00987-UNA (D. Del., Aug 9., 2024) (alleging that AI voice cloning company violated publicity rights of voice-over actors). The case recently settled without any court determination. *See* Joint Stipulation of Voluntary Dismissal with Prejudice, Vacker v. Eleven Labs, No. 1:24-cv-00987-UNA (D. Del., Nov. 5, 2025).

56. *See, e.g.*, Laws v. Sony Music Ent., Inc., 448 F.3d 1134 (9th Cir. 2006); Toney v. L'Oreal USA, Inc., 406 F.3d 905 (7th Cir. 2005); Ahn v. Midway Mfg. Co, 965 F. Supp. 1134 (N.D. Ill. 1997); Eriq Gardner, *"Back to the Future II" from a Legal Perspective: Unintentionally Visionary*, HOLLYWOOD REP. (Oct. 21, 2015) https://www.hollywoodreporter.com/business/business-news/back-future-ii-a-legal-833705/ [https://web.archive.org/web/20260207030629/https://www.hollywoodreporter.com/business/business-news/back-future-ii-a-legal-833705/] (describing film's use of old and new footage to make it seem like same actor appeared in second *Back to the Future* movie when he did not); *see also* ROTHMAN, *supra* note 18, at 167–77 (giving examples of these conflicts).

law allowed such licensing by the copyright holder without requiring additional permission of the singer and preempted her state publicity claim.[57]

Other cases have also allowed the reuse of actors and singers' performances on the basis of copyright law. For example, in *Ahn v. Midway Manufacturing*, a case from the 1990s, a district court held that martial artists and actors' performances used in one video game edition could be used in another version of the video game without requiring additional consent by the performers.[58] One of the few cases to swing the other direction involved the band No Doubt's lawsuit against Activision, which allowed a right of publicity claim to proceed in spite of the authorization to have digital versions of the band appear and sing in a video game because the uses exceeded the boundaries of the relevant contracts.[59]

Copyright law could make reuses of copyrighted material in deepfakes "authorized" even though they would be created without the specific approval of the person depicted and in ways that might deceive the public. This explains why the major actors' union, SAG-AFTRA, is worried about using copyright law in this way and about the scope of prior employment contracts, which were drafted and signed before the world of digital replicas in which we are currently living.[60] It is an open question whether digital replicas are copyrightable. If they are, this may further risk a person agreeing to the capture of their digital replica in one context and then losing control over what is done with this replica; a copyright holder is authorized to reuse the copyrighted material in future works and such reuses are often (though not always) allowed on the basis of copyright law preempting state right of publicity claims.[61]

Federal laws addressing deepfakes could combat the preemptive effect of copyright law in such instances, but if the laws are drafted in ways that also empower third parties to deceptively authorize deepfakes this will worsen the problem of deceptively-authorized deepfakes. Recent proposed legislation at the federal level and some bills passed at the state level explicitly create a digital replica right—originating in the underlying person—but allow someone else to either own these rights or have long-term and largely unrestricted licenses to use them.[62] Some bills explicitly allow "authorized representatives" to approve uses of a person's digital replica without

57. *Laws*, 448 F.3d at 1145–46. The court suggested that the plaintiff Laws might have a contract claim against the original record label, but since the case involved a licensee that issue was not before the court.

58. *Ahn*, 965 F. Supp. at 1140.

59. *See* No Doubt v. Activision Publ'g, 702 F. Supp. 2d 1139 (C.D. Cal. 2010).

60. Press Release, SAG-AFTRA, Gov. Newsom Signs Union-Championed A.I. Bills at SAG-AFTRA Plaza (Sep. 17, 2024), https://www.sagaftra.org/gov-newsom-signs-union-championed-ai-bills-sag-aftra-plaza.

61. *See* Jennifer E. Rothman, Donald C. Brace Lecture, *Copyrighting People*, 72 J. COPYRIGHT SOC'Y 1, 7–9, 17, 26–33 (2025).

62. NO FAKES Act of 2025, *supra* note 6; No AI FRAUD Act of 2024, *supra* note 48; TENN. CODE ANN. § 47-25-1101 *et seq.* (2024) (replacing in its entirety prior publicity statute with the ELVIS Act, Tenn. H.B. 2091, passed in 2024, largely to address digital replicas); CAL. LAB. CODE § 927 (West 2024) (added in 2024 by A.B. 2602, 2023–2024 Leg., Reg. Sess. (Cal. 2024)) (allowing the "creation and use of a digital replica of a person's voice or likeness in place of work the individual would otherwise have performed" without requiring approval of specific uses of replica by person depicted).

requiring any consultation with the person as to how the replica of them is used.[63] Even if legislation limits the duration of such licenses, the digital replicas will likely be able to be reused after their expiration. If the digital replicas are copyrightable or contained within copyrighted works, copyright law expressly allows such continued uses and notably also allows new derivative works to be created from them.[64] There is also a danger that parents could authorize deepfakes of their children for their own financial profits. These too are technically "authorized" in a legal sense, but also unauthorized in the sense that the minors themselves may not have had any say in how they are depicted.[65]

The question of deceptively-authorized deepfakes also arises with somewhat different implications in the context of the dead. Obviously, we cannot authorize uses from the grave, so in one sense no uses of deepfakes of dead people are authorized. However, if we create postmortem rights in the identities of the dead as some state laws do, and as several federal bills have proposed, then uses of the dead could be authorized by their estates or by corporations that own or control their identities.

In short, deceptively-authorized deepfakes cause the same potential harms as unauthorized deepfakes both to the underlying person (if alive) and to the public at large if the fakes are deceptive as to their authenticity or sponsorship. Accordingly, legislation targeting deepfakes needs to address such deceptively-authorized deepfakes, rather than incentivize them. In such instances, we might want to consider who should have the authority to agree to such uses and for how long, as well as whether we should follow the decedents' wishes, if known.[66]

D. FICTIONAL DEEPFAKES

The final category of deepfakes does not involve fakes of real people—instead, they are entirely AI-generated. I dub these *fictional deepfakes*. Of course, all deepfakes are fictional in the broader sense of depicting something that is not real, even though the images or sounds appear authentic. My point here by calling them "fictional" is to emphasize that those depicted are themselves fictional constructs rather than depictions of real, identifiable natural persons. The "people" who are depicted in these

63. NO FAKES Act of 2025, *supra* note 6; *see also* TENN. CODE ANN. §§ 47-25-1103, 47-25-1106(f) (allowing rights to a person's voice or likeness to be transferred to others without limitation and allowing those holding contracts with "recording artist[s]" or that hold licenses for "sound recordings" to enforce rights in another person's voice and likeness); CAL. LAB. CODE § 927 (West 2024) (allowing union to authorize uses of members' digital replicas without consultation with specific person depicted).

64. *See, e.g.*, Laws v. Sony Music Ent., Inc., 448 F.3d 1134 (9th Cir. 2006); *see also supra* notes 56-60 and accompanying text.

65. The NO FAKES Act of 2025, *supra* note 6, is one of the few proposed bills to consider additional protections for minors, notably including court review of such licenses. Even with this added layer of protection, the bill does not address the reuse of digital replicas of minors created during the licensing term after the expiration of the licensing periods, and it allows authorized representatives (perhaps authorized by a minor's parent or guardian) to continue to authorize uses even after a minor has turned eighteen.

66. *See* Allen & Rothman, *supra* note 19, at 333-49 (considering appropriate limits on who can bring postmortem claims and how long such rights should persist, as well as supporting the decedent's preferences in some instances).

deepfakes do *not* exist. Consider the recent coverage of the AI-generated actor Tilly Norwood, who is shown speaking and emoting in short clips in realistic ways,[67] or AI-generated fashion models, or AI-generated songs and music filling Spotify's playlists.[68]

The harms that flow from deepfakes that depict entirely synthetic creations are quite different from deepfakes that depict real and identifiable individuals. Fictional deepfakes might deceive the public into thinking they depict real people, but there is not an underlying person who could suffer harms from such a use. Such fictional deepfakes, however, should still be regulated when they are likely to deceive the public into thinking they depict authentic people. In such instances, the same harms to the public can flow from these deceptive fakes of fictional constructs as do from depictions of real people. If, however, these fictional deepfakes are disclosed or obvious, then they should generally be allowed.

Even though deceptive fictional deepfakes should be regulated, the government should allow room for technological disruption and should not be in the business of determining our preferred diet of music and entertainment. Just as actors must tolerate a reduction in on-screen acting jobs because of the success of reality television, animated shows, and sports broadcasts, actors (and recording artists and models) will

---

67. Gene Maddaus, *Tilly Norwood Creator on Hollywood Backlash Creating Jobs and Full AI Movies: "I Don't Think" People Will "Know the Difference,"* VARIETY (Nov. 11, 2025), https://variety.com/2025/film/news/tilly-norwood-creator-eline-van-der-velden-ai-actress-1236574125/ [https://web.archive.org/web/20251229142308/https://variety.com/2025/film/news/tilly-norwood-creator-eline-van-der-velden-ai-actress-1236574125/]; Leo Barraclough, *AI Actress Tilly Norwood Debuts at Zurich Summit as Industry Grapples with Emerging Tech: We Want Her "to be the Next Scarlett Johansson,"* VARIETY (Sep. 28, 2025), https://variety.com/2025/film/global/ai-actress-tilly-norwood-talent-agents-zurich-summit-1236533454/ [https://web.archive.org/web/20260205222845/https://variety.com/2025/film/global/ai-actress-tilly-norwood-talent-agents-zurich-summit-1236533454/].

68. *See* Ali Rogan, *AI-Generated Models Shake Up the Fashion Industry and Raise Concerns* (Aug. 16, 2025), PBS NEWS, https://www.pbs.org/newshour/show/ai-generated-models-shake-up-the-fashion-industry-and-raise-concerns [https://web.archive.org/web/20260205223934/https://www.pbs.org/newshour/show/ai-generated-models-shake-up-the-fashion-industry-and-raise-concerns]; SERAPHINNE VALORA, https://www.seraphinnevallora.com/ [https://web.archive.org/web/20260205224009/https://www.seraphinnevallora.com/] (describing AI-focused marketing agency); Sarah Perez, *Spotify to Label AI Music, Filter Spam and More in AI Policy Change*, TECHCRUNCH (Sept. 25, 2025), https://techcrunch.com/2025/09/25/spotify-updates-ai-policy-to-label-tracks-cut-down-on-spam/ [https://web.archive.org/web/2/https://techcrunch.com/2025/09/25/spotify-updates-ai-policy-to-label-tracks-cut-down-on-spam/]; Brian Hiatt, *AI "Band" The Velvet Sundown Confirm They're AI—and a "Provocation,"* ROLLING STONE (July 5, 2025), https://www.rollingstone.com/music/music-features/ai-band-the-velvet-sundown-confirm-ai-1235379354/ [https://web.archive.org/web/20260205224253/https://www.rollingstone.com/music/music-features/ai-band-the-velvet-sundown-confirm-ai-1235379354/]; *see also* Ethan Millman, *Spotify to Develop AI Music Products in Partnership with Major Record Labels*, HOLLYWOOD REP. (Oct. 16, 2025), https://www.hollywoodreporter.com/music/music-industry-news/spotify-ai-music-partnership-with-record-labels-1236402698/ [https://web.archive.org/web/20260205224338/https://www.hollywoodreporter.com/music/music-industry-news/spotify-ai-music-partnership-with-record-labels-1236402698/] (describing partnership with Universal Music Group, Sony Music Group, and Warner Music group to develop AI music products).

need to tolerate some computer-generated performances.[69] Unions may want to use collective bargaining agreements to require or maximize uses of real actors, but the law should not mandate this.

With that said, there may be questions as to whether the training of some of the AI-generated characters violates the rights of the real people whose performances or identities were used to train these AI "actors."[70] Additionally, if AI outputs too closely resemble a real person, there could also be potential liability under some of the claims that I consider next in Part IV.

* * *

Along with the parade of horribles that flows from deepfake and generative AI technology there is also much to celebrate about this technology. Deepfake technology can improve (and reduce the costs) of visual effects and enhance storytelling and art. The same technology can help those who have lost or are losing the ability to speak to communicate or allow people to use their own voice to speak in foreign languages.[71] The technology can create interactive replicas of deceased loved ones—something that some of us may regard as creepy but others as profound and comforting.[72] The technology can train the police, members of the military, and others using realistic virtual experiences.[73] It can help those with dyslexia and ADHD engage with printed

69. Animated works usually hire voice actors to voice the characters. These jobs are also at risk with AI-generated voice technology.

70. There are allegations that Tilly Norwood was trained on multiple performances of real people. *See* Conor Murray, *SAG-AFTRA Condemns AI "Actress" Tilly Norwood—Joins Critics Emily Blunt, Whoopi Goldberg and More*, FORBES (Sept. 30, 2025), https://www.forbes.com/sites/conormurray/2025/09/30/sag-aftra-condemns-ai-actress-tilly-norwood-joins-critics-emily-blunt-whoopi-goldberg-and-more/ [https://web.archive.org/web/20251202192449/https://www.forbes.com/sites/conormurray/2025/09/30/sag-aftra-condemns-ai-actress-tilly-norwood-joins-critics-emily-blunt-whoopi-goldberg-and-more/].

71. *See* April Dembosky, *People Who Have Lost Their Voices Are Using AI Technology to Regain Them*, NPR: ALL THINGS CONSIDERED (July 22, 2025), https://www.npr.org/2025/07/22/nx-s1-5449081/people-who-have-lost-their-voices-are-using-ai-technology-to-regain-them [https://web.archive.org/web/20260205224513/https://www.npr.org/2025/07/22/nx-s1-5449081/people-who-have-lost-their-voices-are-using-ai-technology-to-regain-them]; ELEVENLABS, https://elevenlabs.io/ [https://web.archive.org/web/20260205224541/https://elevenlabs.io/] (last visited Feb. 20, 2026) (providing translation into more than thirty languages using voice clones).

72. *See* Tharin Pillay, *"The Dead Have Never Been This Talkative": The Rise of AI Resurrection*, TIME (June 27, 2025), https://time.com/7298290/ai-death-grief-memory/ [https://web.archive.org/web/20260205224621/https://time.com/7298290/ai-death-grief-memory/]; *see also* Jake Kanter, *Michael Caine Partners with AI Company ElevenLabs to Clone His Voice*, DEADLINE (Nov. 11, 2025), https://deadline.com/2025/11/michael-caine-elevenlabs-voice-clone-1236613791/ [https://web.archive.org/web/20260205224900/https://deadline.com/2025/11/michael-caine-elevenlabs-voice-clone-1236613791/] (describing ElevenLabs having voice clone rights or licenses over a host of dead people, including Rock Hudson, John Wayne, Judy Garland, Maya Angelou, Amelia Earhart, and Mark Twain).

73. *See Rethinking Response Part Three: VR Training for Public Safety*, POLICING PROJECT, https://www.policingproject.org/rethinking-response-articles/2025/5/8/part-two-body-worn-camera-analytics-e3zg9-zlhxw [https://web.archive.org/web/20260205224812/https://www.policingproject.org/rethinking-response-

material that would otherwise be difficult to access by having familiar voices read the texts out loud.[74]

We cannot and should not legislate our way out of all market and labor disruptions caused by new technology, but we can require that deepfakes be authorized by the person depicted and that they not deceive the public into thinking the fakes are authentic. With this taxonomy and these benchmarks in hand we can now consider the legal landscape in the United States.

IV. THE LEGAL LANDSCAPE IN THE UNITED STATES

The recent vintage of the term deepfakes (and similar terms, like digital replicas) may explain why calls for new laws to address them have overlooked the many longstanding laws that already address unauthorized uses of a person's identity and that likely apply to unauthorized deepfakes. The concerns expressed in the opening demonstration of today's symposium raised the question of whether Taylor Swift would have claims for an unauthorized deepfake of her. The answer is she would have many potential claims. This is a different question than whether a nonprofit creation of such a deepfake for educational and demonstrative purposes would nevertheless be allowable where abundant efforts were made to highlight that there was no affiliation, participation, or sponsorship by Swift. Such a use would likely either fall outside the scope of some claims or be protected by the First Amendment.

Let's consider as a starting point some of the many laws currently on the books that would give claims to those depicted in deepfakes, whether they are megastars or not. My focus will be primarily, though not exclusively, on claims that could be brought by the real, natural person depicted.

A. STATE LAWS

1. The Right of Publicity/Appropriation Tort

At the heart of the legal regime in the United States that protects identity rights is the right of publicity—a state law, which is also sometimes part of or synonymous with

articles/2025/5/8/part-two-body-worn-camera-analytics-e3zg9-zlhxw] (last visited Feb. 21, 2026); Daniel Coates, *"Realism of the Scenario": Lowell Police Rolls Out Immersive, Virtual Reality Training Machine*, BOS. 25 NEWS (Nov. 6, 2025), https://www.boston25news.com/news/local/realism-scenario-lowell-police-rolls-out-immersive-virtual-reality-training-machine/ZDCVFUAADZFKXNCUE4I56SWZ2A/ [https://web.archive.org/web/20260205225109/https://www.boston25news.com/news/local/realism-scenario-lowell-police-rolls-out-immersive-virtual-reality-training-machine/ZDCVFUAADZFKXNCUE4I56SWZ2A/].

74. *See* John Boitnott, *This Immigrant Founder Taught Himself English—Then Made an App that Helps Others with His Disability (and Speed Readers)*, INC. (Aug. 29, 2017), https://www.inc.com/john-boitnott/how-one-founder-turned-his-dyslexia-into-an-app-th.html [https://web.archive.org/web/20251204113902/https://www.inc.com/john-boitnott/how-one-founder-turned-his-dyslexia-into-an-app-th.html].

privacy laws focused on the misappropriation of a person's identity.[75] In broad strokes, right of publicity laws and the appropriation tort provide civil liability (and sometimes criminal penalties) for unauthorized uses of a person's identity.[76] Liability can arise for uses of a person's name, likeness, voice, or performance, as well as for uses of other indicia of a person's identity.[77] These laws date to the early 1900s and have long protected both commercial and personal interests, and both the famous and the ordinary.[78] Even though these laws are more than one hundred years old, they apply without regard to the technology employed and therefore provide claims in the context of the deepfakes of today.

Almost every state provides either a common law or statutory state law claim that protects against unauthorized uses of a person's name, likeness, voice, or other indicia of identity, and no state has rejected such a right for the living.[79] The boundaries differ in some respects but many states track broad protections provided by the common law appropriation tort (even when adopted by statute). I will provide a few illustrative examples, focusing on the three states that are currently the most active in right of publicity litigation or legislation. First, California—California actually has three or four general right of publicity laws depending on how you count them. The state's common law right of publicity, which should be understood as synonymous with its privacy-based appropriation tort, provides that: "To establish a common law claim a plaintiff must prove: (1) the defendant used the plaintiff's identity; (2) the appropriation was for defendant's advantage, commercially or otherwise; (3) lack of consent; and (4) resulting injury."[80] This law covers unauthorized uses of a person's image or voice in deepfakes

75. *See, e.g.*, Zacchini v. Scripps-Howard Broad. Co., 351 N.E.2d 454, 458–60 (Ohio 1976) (explaining that Ohio's right of privacy encompasses a claim for the "appropriation of a plaintiff's name and likeness" and that "this aspect of privacy" is termed "the right of publicity"), *rev'd on other grounds*, 433 U.S. at 565–66 (understanding plaintiff's state right of privacy claim as one for the violation of the right of publicity); Prima v. Darden Rests., Inc., 78 F. Supp. 2d 337, 346 (D.N.J. 2000) ("Louisiana law . . . does not expressly provide for a right of publicity. Rather, courts in Louisiana have interpreted Louisiana's right of privacy to protect a person's name or likeness from commercial exploitation.") (citing Prudhomme v. Procter & Gamble Co., 800 F. Supp. 390, 396 (E.D. La. 1992)); Brinkley v. Casablancas, 438 N.Y.S.2d. 1004, 1012 (N.Y. App. Div. 1981) (concluding that New York's privacy statute is the state's "right of publicity"); *see also* RESTATEMENT (SECOND) OF TORTS § 652C (1977) (treating the privacy-based appropriation tort and the right of publicity as identical); ROTHMAN, *supra* note 18, at 11–86; Post & Rothman, *supra* note 18, at 93–95.

76. *See* RESTATEMENT (SECOND) OF TORTS § 652C (1977); Eastwood v. Superior Court, 149 Cal. App.3d 409, 416–17 (1983); N.Y. CIV. RIGHTS LAW §§ 50–51; *see also* ROTHMAN, *supra* note 18; ROTHMAN'S ROADMAP TO THE RIGHT OF PUBLICITY, https://rightofpublicityroadmap.com/ https://web.archive.org/web/20260226131359/https://rightofpublicityroadmap.com/ (last visited Mar. 7, 2026) [hereinafter "ROTHMAN'S ROADMAP"].

77. *See, e.g.*, *Zacchini*, 433 U.S. 562; Jordan v. Jewel Food Stores, Inc., 743 F.3d 509 (7th Cir. 2014); Abdul-Jabbar v. General Motors Corp., 85 F.3d 407 (9th Cir. 1996); Waits v. Frito-Lay, Inc., 978 F.2d 1093 (9th Cir. 1992); White v. Samsung Elecs. Am., 971 F.2d 1395 (9th Cir. 1992); Midler v. Ford Motor Co., 849 F.2d 460 (9th Cir. 1988); Motschenbacher v. R.J. Reynolds Tobacco Co., 498 F.2d 821 (9th Cir. 1974); *see also* ROTHMAN, *supra* note 18, 88–96 (describing history and sweep of right of publicity laws).

78. *See* ROTHMAN, *supra* note 18, at 27–44.

79. *See* ROTHMAN'S ROADMAP, *supra* note 76. States, such as New York, that initially rejected such a claim under their common law have subsequently adopted similar provisions by statute.

80. *Eastwood*, 149 Cal. App. 3d at 416–17.

and does so regardless of whether the person depicted has a commercially valuable identity and regardless of whether the use was a commercial one.

California's statutory right of privacy/publicity for the living, located in Civil Code § 3344, also covers deepfakes though with a narrower scope—limited to uses "on or in products, merchandise, or goods, or for purposes of advertising or selling, or soliciting purchases of, products, merchandise, goods or services."[81] This law was passed to extend statutory damages and attorney's fees to people who were not celebrities to make such claims viable for ordinary people.[82] California's right of publicity for the dead was revised in 2025 to explicitly address digital replicas. The amendments to the statute remove the statutory exemption for uses in "audiovisual work[s]" in the context of digital replicas that depict "deceased personalit[ies]."[83] California's postmortem statute does not apply to everyone—something that I have criticized elsewhere—but only applies to well-known individuals who died with "commercial value" at the time of their death.[84]

New York's statutory privacy and publicity laws also cover deepfakes but might not cover noncommercial deepfakes because it is limited to claims that arise out of uses for advertising or trade purposes.[85] *Lehrman v. Lovo* is one of the first cases to hold that New York's statutory privacy/publicity law bars unauthorized voice clones.[86] New York's postmortem publicity right, which was added in 2020, explicitly covers deepfakes or digital replicas of dead people. Notably, the scope of the postmortem provision is very narrow, only applying to "deceased performers."[87]

Tennessee's 2024 overhaul of its right of publicity laws was expressly done to address concerns over deepfakes.[88] It added numerous provisions, including adding "voice" to its statutory bar on unauthorized uses of a person's identity, which now explicitly covers the unauthorized use of a person's "name, photograph, voice, or

81. CAL. CIV. CODE § 3344 (West 2024).

82. California's statutory right of publicity for the living, which is now frequently used to protect the commercial value of identity, was originally passed under the moniker of "privacy" and created to provide ordinary citizens whose identity lacked commercial value the opportunity to obtain statutory damages. Act of Nov. 22, 1971, ch. 1595, 1971 Cal. Stat. 3426 (*codified at* CAL. CIV. CODE § 3344 (West 2024)); ROTHMAN, *supra* note 18, at 208 n.40; Letter from John Vasconcellos, Member, Cal. State Assembly, to Ronald Reagan, Governor of Cal. (Nov. 10, 1971) (on file with the Cal. State Archives, Governor's Chaptered Bill File).

83. CAL. CIV. CODE § 3344.1 (West 2025). This postmortem statute was originally passed in 1984 (as CAL. CIV. CODE § 990) and long had exemptions for uses in various expressive and artistic contexts, notably including audiovisual works and accordingly there were concerns that digital replicas of the dead might be exempted from liability—something that was not the case with the right for the living which does not have similar exemptions either at common law or under the statute.

84. CAL. CIV. CODE § 3344.1; Allen & Rothman, *supra* note 19, at 335–36.

85. N.Y. CIV. RIGHTS LAW §§ 50–51.

86. Lehrman v. Lovo, Inc., 790 F. Supp. 3d 348 (S.D.N.Y. 2025) (allowing right of publicity/privacy claims to proceed beyond the motion to dismiss stage under N.Y. CIV. RIGHTS LAW §§ 50–51).

87. N.Y. CIV. RIGHTS LAW § 50-f. Deceased performers are narrowly defined as "a deceased natural person domiciled in [the State of New York] at the time of death who, for gain or livelihood, was regularly engaged in acting, singing, dancing, or playing a musical instrument." *Id.* § 50-f(1)(a).

88. TENN. CODE ANN. §§ 47-25-1101–1108 (2024) (replacing in its entirety the prior publicity statute with the ELVIS Act, passed in 2024).

likeness."[89] The law protects all people regardless of whether they commercialize their identities. The statute creates liability when unauthorized uses of a person's name, voice, or likeness are "for purposes of advertising products, merchandise, goods, or services, or for purposes of fundraising" or a "person publishes, performs, distributes, transmits, or otherwise makes available to the public an individual's voice or likeness, with knowledge that use of the voice or likeness was not authorized."[90] The statute also bars the distribution or making available of "an algorithm, software, tool, or other technology, service, or device, the primary purpose or function of [which] is the production of a particular, identifiable individual's photograph, voice, or likeness, with knowledge that distributing, transmitting, or otherwise making available the photograph, voice, or likeness was not authorized."[91] Thus, the law not only regulates the making of deepfakes but also the software that facilitates their creation and creates liability for knowing dissemination of deepfakes. It also makes clear that although an audiovisual work can represent a real person, it cannot do so in a way that is "intended to create, and does create, the false impression that the work is an authentic recording in which the individual participated."[92] Tennessee may also have a broader common law right of publicity.[93]

In short, most state right of publicity and privacy laws cover deepfakes,[94] but there are some reasons we might want more protection against deepfakes than these laws provide. First off, and importantly, most right of publicity laws allow deceptive uses so long as they are authorized by the publicity holder. Right of publicity laws therefore do not address authorized deepfakes that nevertheless deceive the public. Second, some of these right of publicity laws explicitly allow ownership or control by someone other than the person depicted.[95] Thus, they allow deceptively-authorized deepfakes—which, as I have observed, cause all of the same harms as unauthorized deepfakes. If such control by others is allowed, these laws could amplify rather than limit the deception of the public and also harm the individuals depicted.

89. TENN. CODE ANN. § 47-25-1105 (2024).

90. TENN. CODE ANN. §§ 47-25-1105(a)(1) & (2) (2024).

91. *Id.* § 47-25-1105(a)(3).

92. TENN. CODE ANN. § 47-25-1107(a)(3) (2024).

93. There is an open question of whether Tennessee also has a common law right of publicity/appropriation right or whether this right was "supplant[ed]" by the statute. *See* Marshall v. ESPN, Inc., 111 F. Supp. 3d 815, 824 (M.D. Tenn. 2015). No state court in Tennessee has ruled that the common law is preempted by the statute. *See, e.g.*, State *ex rel.* Elvis Presley Int'l Mem'l Found. v. Crowell, 733 S.W.2d 89 (Tenn. Ct. App. 1987) (indicating that the state recognizes both a common law and statutory right of publicity and privacy).

94. For some illustrative examples of publicity rights in other states that would cover deepfakes, *see, e.g.*, Ventura v. Titan Sports, Inc., 65 F.3d 725 (8th Cir. 1995) (recognizing common law publicity rights in Minnesota); Lake v. Wal-Mart Stores, Inc., 582 N.W.2d 231 (Minn. 1998) (recognizing appropriation tort in Minnesota); Doe v. TCI Cablevision, 110 S.W.3d 363 (Mo. 2003) (recognizing both a common law right of publicity and an appropriation tort in Missouri, differentiated solely over whether use is for commercial advantage or any advantage); Hepp v. Facebook, 14 F.4th 204 (3d Cir. 2021) (recognizing both a statutory and common law right of publicity in Pennsylvania); Vogel v. W. T. Grant Co., 327 A.2d 133 (Pa. 1974) (recognizing common law appropriation tort in Pennsylvania).

95. *See, e.g.*, ALA. CODE § 6-5-771 (2024); CAL. LAB. CODE § 927 (2024); LA. REV. STAT. § 51-470.3; TENN. CODE ANN. §§ 47-25-1103, 47-25-1106 (2024).

Third, some state publicity laws limit claims to the context of commercial uses or for commercial advantage. This may make it difficult to make out a right of publicity case for deepfakes made by noncommercial entities and that are circulated for free in circumstances that are not meant to boost a person's commercial footprint. A number (albeit a minority) of states also limit who can bring claims to those with commercially valuable identities—primarily in the context of postmortem rights—which will make it hard for everyone to make out a publicity/privacy claim. Even in the vast majority of jurisdictions that allow ordinary people to bring publicity/appropriation claims, damages may be hard to prove and not all states provide statutory damages or fee-shifting provisions which would make such claims viable.[96]

Additionally, as discussed, copyright law may preempt state publicity claims if they arise out of deepfakes created by an authorized owner or licensee of copyrighted works used to create the fakes. There is also a circuit split on whether a federal law, the Communications Decency Act Section 230, may limit platform liability for right of publicity claims arising out of uses by third parties.[97] If platforms are immunized from such claims, they will have less of an incentive to remove unauthorized deepfakes, and plaintiffs will not be able to recover damages from the platforms.

Finally, the simple fact that we have fifty different state laws to navigate poses some challenges given the likely dissemination of works across the country and some variability across the states in what rights are extended.[98] Although most states recognize some form of the appropriation tort which is largely the same across jurisdictions, there is great variability in the statutory rights, especially with regard to the treatment of the dead.

---

96. Some states realizing this problem have added statutory damages and fee-shifting provisions to ensure that ordinary people are adequately protected and that there is a sufficient incentive not to use a person's identity without regard to their commercial value. California, for example, has long had statutory damages in its statutory publicity/privacy law, as well as a fee-shifting provision. CAL. CIV. CODE § 3344 (2025); *see also supra* note 82. California also recently passed a law increasing statutory damages to $250,000 for the "malicious" circulation of unauthorized intimate images, including deepfakes. *See* A.B. 621, 2025 Gen. Assemb., Reg. Sess. (Ca. 2025) (signed by governor Oct. 13, 2025); *see also, e.g.*, ALA. CODE § 6-5-774 (2024) (providing $5,000 statutory damages).

97. *Compare* Hepp v. Facebook, 14 F.4th 204 (3d Cir. 2021) (holding that § 230 does not bar state right of publicity claim from proceeding against Facebook) *with* Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102 (9th Cir. 2007) (holding that right of publicity claim could not proceed because of § 230 immunity); *see also* Jennifer E. Rothman, *Third Circuit Holds that Newscaster's Right of Publicity Claim Can Proceed Against Facebook*, ROTHMAN'S ROADMAP (Sep. 28, 2021), https://rightofpublicityroadmap.com/news_commentary/third-circuit-holds-that-newscasters-right-of-publicity-claim-can-proceed-against-facebook/ [https://web.archive.org/web/20260208092309/https://rightofpublicityroadmap.com/news_commentary/third-circuit-holds-that-newscasters-right-of-publicity-claim-can-proceed-against-facebook/]. Some states have tried to address this issue by expressly stating in statutes that the statutory publicity right (or related law) is a form of intellectual property for purposes of § 230. State statutes cannot, however, determine what is meant by a federal law. Nevertheless, if the federal law includes state intellectual property laws, then state designations may affect whether the § 230 exemption applies.

98. I note that to address some of this variability, I have been charged as the Reporter of the Uniform Law Commission's Study of Name, Image, and Likeness Rights to consider the possibility of drafting a uniform law to create greater uniformity across state laws.

2. Other State Laws Targeting Identity Rights

Right of publicity laws are not the only state laws that apply to deepfakes. Many other state statutes specifically target identity rights and cover deepfakes. These include intimate image laws,[99] biometric privacy laws,[100] impersonation laws (sometimes referred to as catfishing laws),[101] student-athlete NIL (name, image, likeness) laws,[102] digital replica laws, election laws, labor laws, and a host of other AI-specific laws.[103]

---

99. *See, e.g.*, 750 ILL. COMP. STAT. 5/11-23.5 (2024) (criminalizing the "[n]on-consensual dissemination of private sexual images"); TEX. PENAL CODE ANN. § 21.16 (West 2024) (same). Some of these laws expressly cover digital replicas or deepfakes. *See infra* note 103.

100. *See, e.g.*, Biometric Information Privacy Act, 740 ILL. COMP. STAT. 14 *et seq.* (2024); TEX. BUS. & COM. CODE ANN. § 503.001 *et seq.* (West 2024) (limiting capturing and use of biometric identifiers).

101. *See, e.g.*, CAL. PENAL CODE § 528.5 (West 2024) (added in 2011); OKLA. STAT. TIT. 21, § 1450 (2024) (added in 2016); H.B. 783, 89th Leg., Reg. Sess. (Tex. 2025) (law in Texas enacted in 2025 that provides civil liability for "online impersonation").

102. Since 2020, more than thirty states have passed new laws (or amendments to existing ones) that address name, image, and likeness rights for student-athletes. These are often referred to as "NIL" rights even though student-athletes are covered by broader publicity rights that already protect their name, image, and likeness rights. The recent surge in the use of NIL as a term of art has obscured the longstanding existence of such rights. Examples of recently passed bills addressing student athletes include: ARK. ANN. CODE § 4-71-1301 *et seq.* (2024) (providing a specific publicity right for student-athletes and providing some limits on this right) (effective date Jan. 1, 2022); 70 OK. STAT. § 820.21 *et seq.* (2025) (extending NIL rights to student athletes and providing some limits on the right) (effective date May 28, 2021, but amended several times including as recently as May 2025). Several bills have been floated in Congress that would extend federal "name, image, and likeness rights" for college athletes but so far none has passed. *See, e.g.*, The Student Athlete Fairness and Enforcement (SAFE) Act, S. 2932, 119th Cong. (introduced Sept. 29, 2025); College Athlete Economic Freedom Act, H.R.4868, 119th Cong. (introduced Aug. 1, 2025); Student Compensation and Opportunity Through Rights and Endorsements (SCORE) Act, H.R. 4312, 119th Cong. (introduced July 10, 2025). Some of these bills are somewhat strange given that they are drafted as if student-athletes would have no claims under current law, which is not the case. Present Trump recently signed an Executive Order seeking to address the Wild West atmosphere of today's college sports endorsement and payment regime. *See* Exec. Order No. 14,322, 90 Fed. Reg. 35,821 (July 24, 2025) ("Saving College Sports").

103. Some of these laws are general purpose statutes targeting deepfakes across contexts. *See, e.g.*, H.B. 1432, 2024 Sess. (N.H. 2024) (creating a felony if a person knowingly creates, distributes, or presents a deepfake for the purpose of causing financial or reputational harm to another) (effective Jan. 1, 2025). Some laws, while written broadly, have particularly targeted concerns raised by the recording industry. *See, e.g.*, A.B. 1836, 2023–24 Reg. Sess. (Cal. 2024) (enacting protections against digital replicas in amendments to postmortem right of publicity statute, CAL. CIV. CODE § 3344.1 (West 2024)); TENN. CODE ANN. § 47-25-1101 (2024) (adopting ELVIS Act in 2024); LA. REV. STAT. § 51-470.1 et seq. (2025) (adopting broad statutory right of publicity and including consideration of digital replicas).

Other digital replica laws particularly target pornographic contexts or uses that depict intimate parts of a person's body (whether computer-generated or otherwise). *See, e.g.*, CAL. CIV. CODE § 1708.86 (West 2024) (creating civil liability when "an individual . . . appears, as a result of digitization, to be giving a [sexual] performance they did not actually perform or to be performing in an altered depiction") (amended on Oct. 13, 2025 to expand scope to include nudifying, better protect minors, and raise statutory damages to $250,000, see A.B. 621, 2025 Gen. Assemb., Reg. Sess. (Ca. 2025)); 750 ILL. COMP. STAT. 5/11-23.7 (2024) (criminalizing "[n]on-consensual dissemination of sexually explicit digitized depictions") (effective Jan.1, 2025); N.Y. CIV. RIGHTS LAW § 52-C (McKinney 2024) (providing a civil action for circulation of nonconsensual explicit depictions).

Other state laws have targeted deepfakes in the context of elections. *See, e.g.*, CAL. ELEC. CODE § 20010 (West 2024) (making actionable the distribution, "with actual malice," of "materially deceptive" deepfakes "of [a] candidate with the intent to injure the candidate's reputation or to deceive a voter into

Hundreds of these laws have passed in the last five years, making it impossible to encapsulate them here.[104] It is worth highlighting, however, that some of these laws face the same challenges of publicity laws in that many do not focus on regulating public deception and some allow for parties other than the person depicted to authorize deepfakes of that person.

3. General State Laws that Apply to Deepfakes

There are a host of other state-based claims relevant to identity rights that may also apply in some instances of deepfakes. Many of these claims are likely to be successful in the context of regulating deepfakes, particularly claims for the intentional and negligent infliction of emotional distress, defamation, the false light tort, and, as I will discuss further below, unfair competition and trademark laws.

* * *

In short, there are many state laws that already apply to deepfakes, but the sheer number of them and their variations across jurisdictions may be difficult to navigate both for potential plaintiffs and for creators that are trying to stay within the boundaries of the law across jurisdictions.

B. FEDERAL LAWS

There are a number of federal laws, including trademark and copyright laws, that also apply to deepfakes, but not in every instance or for every possible plaintiff. I will consider here both existing laws and some proposed legislation.

Although people are not copyrightable, it is an open question whether a person's digital replica or voice clone could be.[105] Regardless, people can be captured in copyrightable works, and if they retain the copyright in these works or obtain rights (or licenses) to them, then they can use copyright law to restrict unauthorized uses of their images, likeness, performances, and voices that have been captured in these works

---

voting for or against the candidate"); TEX. ELEC. CODE § 255.004(d) (West 2024) (criminalizing creating and distributing "a video, created with the intent to deceive, that appears to depict a real person performing an action that did not occur in reality," when the creation and distribution is "with intent to injure a candidate or influence the result of an election").

Several states have adopted new laws or amended their employment and labor laws to regulate contracts for the creation and use of digital replicas. *See, e.g.*, CAL. LAB. CODE § 927 (West 2024); Public Act 104-0282 (Ill. 2024); N.Y. GEN. OBLIG. LAW § 5-302 (McKinney 2024). New York recently adopted a law requiring advertisers to disclose when a digitally created "synthetic performer" is used in advertisements. *See* S.8420-A/A.8887-B, 2025–2026 Reg. Sess. (N.Y. 2025) (signed into law December 11, 2025).

104. *See, e.g., supra* notes 99–103; *see also* Thomas E. Kadri & Sonja R. West, *Deepfake Torts: Emerging Torts Frameworks in the U.S. Deepfake Regulation*, 18 J. TORT L. 515, 518–19 (2025) (documenting 122 new deepfake laws passed during the period between January 2023 and May 2025, and 344 proposed laws related to regulating deepfakes).

105. *See* Rothman, *supra* note 61.

and the creation of derivatives based on them.[106] There are a number of cases currently being litigated to determine the ability of copyright law to limit the creation of digital replicas and voice clones, but they are still in the early stages or have settled, leaving the legal questions unanswered.[107]

Unfair competition and trademark laws both at the state and federal level also protect against unauthorized uses of a person's identity, most often uses of a person's name or likeness, but also voice.[108] These laws protect individuals by restricting the unauthorized use of a person's identity as a mark (or otherwise) to falsely indicate source, endorsement, or sponsorship and in some instances to prevent the dilution of a famous "personal mark."[109] Such claims are usually limited to those that arise from using the identity of a person who sells goods or services associated with their identity. State and federal consumer protection laws and regulations also limit unauthorized uses of a person's identity, as do laws prohibiting fraud and false advertising.

In contrast to right of publicity claims, all of these claims (other than dilution) require a demonstration of likely confusion of consumers. These unfair competition

106. *See, e.g.*, Balsley v. LFP, Inc., 691 F.3d 747 (6th Cir. 2012) (using copyright to limit circulation of plaintiff in wet t-shirt contest); Monge v. Maya Magazines, Inc., 688 F.3d 1164, 1184 (9th Cir. 2012) (copyright limited unauthorized circulation of wedding photos); Michaels v. Internet Ent. Grp., Inc., 5 F. Supp. 2d 823 (C.D. Cal. 1998) (using copyright to limit circulation of sex tape); *see also* Shyamkrishna Balganesh, *Privative Copyright*, 73 VAND. L. REV. 1, 4–5 (2020); Rothman, *supra* note 61, at 18–24.

107. *See, e.g.*, Lerhman v. Lovo, Inc., 790 F. Supp. 3d 348 (S.D.N.Y. 2025) (allowing right of publicity claims to proceed and leave to amend copyright claims arising out of unauthorized use of plaintiffs' voices in AI-generated voice clones); Complaint at 23–25, UMG Recordings, Inc. v. Suno, Inc., No. 24-11611, 2025 WL 3524289 (D. Mass., filed June 24, 2024) (contending that AI-generated songs violate copyright in recordings in part on basis of use of performers' voices); Complaint, UMG Recordings, Inc. v. Uncharted Labs, Inc., No. 24-04777, 2025 WL 1047517 (S.D.N.Y., filed June 25, 2024) (objecting to use of recording artists' "vocal style" from copyrighted recordings in AI-generated output, including outputs similar to Mariah Carey, Bruce Springsteen, and Frank Sinatra); *cf.* Order Denying Motion to Dismiss, Concord Music Grp., Inc. v. Anthropic PBC, No. 24-03811-EKL, (N.D. Cal. Oct. 6, 2025) (allowing copyright claims to proceed arising out of use of song lyrics in training AI models and their outputs). Several of the lawsuits brought by record labels have settled. Notably, the settlements have involved partnerships to create AI-generated music with the defendants rather than to limit their actions. *See, e.g.*, Stipulation of Dismissal, UMG Recordings, Inc. v. Suno, Inc., No. 24-11611 (D. Mass., filed Jan. 28, 2026) (dismissing case brought by Atlantic Records, Rhino Entertainment, Warner Music, and others); Stipulation of Dismissal by Plaintiffs UMG Recordings, Inc., and Capitol Records, LLC, No. 24-04777 (S.D.N.Y., filed Nov. 5, 2025); Stipulation of Dismissal by Plaintiffs Atlantic Recording Corp. et al., LLC, No. 24-04777 (S.D.N.Y., filed Nov. 5, 2025) (dismissing case brought by Warner Music, Rhino Entertainment, and others); *Warner Music Group, Udio Settle Copyright Case, Plan New AI Song Creation Platform*, REUTERS (Nov. 19, 2025), https://www.reuters.com/legal/litigation/warner-music-settles-with-ai-firm-udio-plans-joint-platform-2025-11-19/; Press Release, Warner Music Group, Warner Music Group and Suno Forge Groundbreaking Partnership (Nov. 25, 2025), https://www.wmg.com/news/warner-music-group-and-suno-forge-groundbreaking-partnership [https://perma.cc/T266-KT77].

108. *See* Rothman, *supra* note 19, 1278–79; *see also* Rachel Buckland & Shivani Chelliah, *Much Ado About McConaughey*, ROTHMAN'S ROADMAP (Feb, 27, 2026), https://rightofpublicityroadmap.com/news_commentary/much-ado-about-mcconaughey/ [https://web.archive.org/web/20260323224332/https://rightofpublicityroadmap.com/news_commentary/much-ado-about-mcconaughey/] (analyzing recent press coverage of actor's trademark registrations given long history of personal marks, including their registration at the USPTO).

109. "Personal marks are those that include (or are entirely composed of) the portrait, name, or other indicia of identity of a natural person." Rothman, *supra* note 19, at 1276.

laws could help combat public deception to some degree in the context of deepfakes, but not entirely; the focus of the deception inquiry in these laws is usually on confusion as to source, sponsorship, or affiliation rather than on whether the public is deceived into thinking a deepfake is an authentic recording. Additionally, such laws will usually not provide claims for those whose identities do not have commercial value.

Congress has also recently passed and is considering passing other laws that address certain aspects of name, image, likeness and voice rights.[110] The Take It Down Act recently signed into law by President Trump facilitates the removal from online platforms of unauthorized intimate visual depictions, including AI-generated ones.[111] Of the many other bills being floated to address deepfakes, the NO FAKES Act thus far appears to have the most support of these proposed bills. The title is short for "Nurture Originals, Foster Art, and Keep Entertainment Safe Act." The title itself highlights that this was drafted with the entertainment industries in mind, even though it would apply to everyone. The bill would create a federal digital replica right, including a postmortem right, and create liability for producing, publishing, reproducing, displaying, distributing, transmitting or "otherwise making available to the public" a digital replica.[112] It provides statutory damages ranging from $5,000 to $25,000 for entities that are not online service providers.[113] The bill would allow for some reuse of copyrighted works (including sampling, remixing, and remastering, and reuse of some performances if they are not "fundamentally" or "materially altered").[114]

The bill runs thirty-nine pages—and is still a work-in-progress. In the context of our symposium's topic of deepfakes, I want to therefore focus on how the proposed NO FAKES Act would address the key concerns that we identified arising out of deepfakes that center on authorization and deception. The short answer is that it doesn't do a good job of addressing either concern. First, it says nothing about addressing public deception and may increase the number of deceptive deepfakes by further incentivizing a market for them. Second, the bill doesn't protect against deceptively-authorized deepfakes—remember, those are the deepfakes that might technically be authorized as a legal matter but that are not in fact specifically known or approved by the identity-holder. This is so because the bill allows long-term, broad licensing with insufficient guardrails and allows authorized representatives to enter such licensing agreements without consulting the identity-holder.[115] As the bill is redrafted, deception and

110. *See, e.g.*, Take It Down Act, Pub. L. No. 119-12, 139 Stat. 55 (2025) (codified as amended at 47 U.S.C. §§ 223–223a); NO FAKES Act of 2025, *supra* note 6; *see also* No AI FRAUD Act of 2024, *supra* note 48; PADRA, *supra* note 6.

111. Take It Down Act, *supra* note 110. Notably, a district court has recently called into question the law's effectiveness by suggesting that § 230 of the Communications Decency Act blocks platform liability. Doe v. X Corp., No. 4:25-cv-01282-0, at 11–12 (N.D. Tex. 2026).

112. NO FAKES ACT of 2025, *supra* note 6.

113. *Id.* § 2(e)(4).

114. *Id.* § 2(a)(2).

115. Although the bill requires that licenses identify the uses to be exercised with a "reasonably specific description," this does not require the person's actual knowledge and it is not clear how even this proposed limitation would be interpreted. For example, would a description of the use of a person's digital replica as "in audio-visual works" be "reasonably specific"? Would uses in promotions for a particular brand of soda or

authorization should be at the center of the drafting rather than profit-maximization for third parties and immunization for internet platforms.

* * *

In sum, even though we have many—maybe too many—overlapping and sometimes conflicting laws in the United States that address deepfakes, they frequently do not keep their eye on the key dangers. Often both the laws on the books (and those being proposed) do not adequately focus on whether the deepfakes are authorized by those depicted and/or whether the deepfakes deceive the public as to their authenticity.[116] The current explosion in overlapping laws that all extend rights or claims based on the attributes of a real person have also exacerbated what I elsewhere have dubbed an "identity thicket," further complicating the navigation of the legal landscape covering deepfakes.[117] Ideally, any new federal laws would try to thin out the "identity thicket" rather than worsen it.

V. CONCLUSION

With all of this discussion of the law, it's important to note that law alone cannot combat deepfakes. We will need to work with technologists to build guardrails against the creation and dissemination of unauthorized and deceptive deepfakes. We need tools to prove authenticity and to detect fakes. Laws can be designed to encourage such private sector efforts but at the federal level this has not been the focus. Some states, particularly California, however, have tried to do this and we will have to see how successful these efforts are. Requiring platforms to address deepfakes is also essential since having laws on the books that are hard to or impossible to enforce will have little impact.

We also do not know how the market will react to this era of deepfakes. For example, we may find that people will crave authentic performances, perhaps reviving live theater and preferring DJ-moderated, human-created music to AI-generated playlists filled with fake recordings by fictional artists. Certification regimes that mark things as human-made, among other market-driven approaches, could help support these consumer choices.

In our rush to fix the problem of deepfakes, we should make sure that the law does not worsen the problem by giving legitimacy to deceptively-authorized deepfakes. Nor

apparel be "reasonably specific"? The limitation is essentially useless if both of these examples count. For consideration of these and other concerns with the current draft of the NO FAKES Act, see Rothman, *Still Needs Revision*, *supra* note 47.

116. Addressing these crucial concerns at the same time and across contexts would have the added benefit of reducing the underinclusivity concerns of more targeted approaches. See discussion *supra* note 31 and accompanying text. It could also highlight alternatives to mandated disclosures that may be struck down as compelled speech or that may have limited effectiveness. *See* Nat'l Inst. of Family & Life Advocates v. Becerra, 585 U.S. 755 (2018) (holding that state's notice requirement for crisis pregnancy centers is likely unconstitutional).

117. Rothman, *supra* note 19, at 1273, 1278–89.

should the law overlook the harms that flow from even authorized deepfakes that deceive the public. Unfortunately, too much of the recently proposed and enacted legislation does exactly this. Given that we already have many laws at both the state and federal level that apply to deepfakes, further legislation in this area should be squarely focused on addressing the current gaps in the law, with an eye toward limiting deceptive uses and ensuring that the people depicted truly agreed to the deepfake portrayal.

From Merchandise to Movies: The Rapid—and Potentially Worrisome—Expansion of NILV Regulation to Cover Uses in Expressive Works

Benjamin S. Sheffner*

* Senior Vice President & Associate General Counsel, Law & Policy, Motion Picture Association, Inc. ("MPA"). The views expressed in this article are the author's own and do not necessarily reflect those of the MPA or its members.

INTRODUCTION

Drive down the highway and see your face on a billboard advertising a product, when you had never given permission to use your likeness? Notice your visage used on a piece of merchandise when you had never consented to such use? Hear your voice used in a jingle to plug a product? The law in virtually every state provides a clear remedy for such commercial uses of one's name, image, likeness, or voice: Sue for violation of your right of publicity, or, as termed in some jurisdictions, your right of privacy.[1] But what if your likeness or voice appears in a movie or television show, or a YouTube or TikTok video, not to hawk a product, but to tell a story, or as part of a parody or satire, or to make a political point? Such uses have traditionally remained outside of the ambit of right of publicity law.

Now? It's not so clear. As this Article will detail, there has been a recent sea change in U.S. state law regulating name, image, likeness, and voice ("NILV"). Concerned by technological developments that have enabled the easy creation of "digital replicas" or "deepfakes" of individuals, and fueled by lobbying by the recording industry and representatives of actors and singers, states have, over the past decade, proposed and in some cases enacted numerous new laws regulating the use of NILV outside the advertising/merchandising context, and explicitly in the context of expressive works such as movies and songs.[2] And Congress is considering a bill that, if enacted, would establish an entire new federal intellectual property right covering likeness and voice.[3] The emergence of new forms of harm may justify new laws regulating NILV. But these

1. *See, e.g.*, CAL. CIV. CODE § 3344(a) ("Any person who knowingly uses another's name, voice, signature, photograph, or likeness, in any manner, on or in products, merchandise, or goods, or for purposes of advertising or selling, or soliciting purchases of, products, merchandise, goods or services, without such person's prior consent, or, in the case of a minor, the prior consent of his parent or legal guardian, shall be liable for any damages sustained by the person or persons injured as a result thereof"); N.Y. CIV. RIGHTS LAW § 51 ("Any person whose name, portrait, picture, likeness or voice is used within this state for advertising purposes or for the purposes of trade without the written consent first obtained as above provided may maintain an equitable action in the supreme court of this state against the person, firm or corporation so using such person's name, portrait, picture, likeness or voice, to prevent and restrain the use thereof . . ."). For a comprehensive overview of state right of publicity laws, see JENNIFER E. ROTHMAN, ROTHMAN'S ROADMAP TO THE RIGHT OF PUBLICITY, https://rightofpublicityroadmap.com/ [https://web.archive.org/web/20260216052417/https://rightofpublicityroadmap.com/] (last visited Feb. 23, 2026).

2. *See infra* Part II.

3. *See* Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2025 ("NO FAKES Act"), S. 1367, 119th Cong. (2025).

new laws, which by their very terms regulate the content of speech—which makes them "presumptively unconstitutional"[4]—raise serious questions about whether they are consistent with First Amendment protections for non-commercial speech, and will likely be the subject of litigation challenging their constitutionality in coming decades.

I. EXISTING STATE RIGHT OF PUBLICITY LAWS

Right of publicity is the body of state law prohibiting unauthorized exploitation of an individual's name, image, likeness, or voice, typically for commercial purposes, such as in an advertisement or on merchandise. Today, in virtually every state, an individual has a cause of action if his or her likeness is used without permission on a billboard or in a television advertisement, or on a product like a coffee mug or cereal box. These laws are by their terms technology-neutral; they apply to uses of an individual's NILV whether done using traditional technologies like photography, or via new technologies like digital replicas created by artificial intelligence-powered tools.[5]

Of crucial importance, right of publicity properly understood has no place in the context of non-commercial speech—speech in the form of movies, television programs, books, news articles and broadcasts, songs, etc., often termed "expressive works"[6]—which receives full First Amendment protection.[7] Recognizing the need to prevent right of publicity statutes from encroaching on First Amendment rights, states that have enacted or amended such statutes in the past several decades have routinely included explicit statutory exceptions, known as "expressive-works exemptions," which make clear that this body of law has no application in the context of creative and journalistic works.[8] States have codified these exemptions precisely because they know that, if not properly cabined to commercial uses, right of publicity statutes risk chilling

4. Reed v. Town of Gilbert, Ariz., 576 U.S. 155, 163 (2015).

5. *See* Lohan v. Take-Two Interactive Software, Inc., 31 N.Y.3d 111, 121–22 (2018) (holding that a digital avatar may qualify as a "portrait" under New York right of privacy law).

6. The term "expressive works" refers to works that are "not an advertisement for or endorsement of a product." Comedy III Prods., Inc. v. Gary Saderup, Inc., 25 Cal. 4th 387, 396 (2001).

7. Under well-established First Amendment doctrine, the term "commercial speech" refers to "speech which does no more than propose a commercial transaction." Va. Pharmacy Bd. v. Va. Consumer Council, 425 U.S. 748, 762 (1976) (internal quotation marks omitted). Artistic, literary, journalistic, and similar forms of speech are not "commercial speech"—even if sold for a profit. *See* Joseph Burstyn, Inc. v. Wilson, 343 U.S. 495, 501 (1952) ("It is urged that motion pictures do not fall within the First Amendment's aegis because their production, distribution, and exhibition is a large-scale business conducted for private profit. We cannot agree."); 303 Creative LLC v. Elenis, 600 U.S. 570, 594 (2023) (rejecting argument that speech receives lesser First Amendment protection when sold for a profit or by a corporation: "none of that makes a difference"); Sarver v. Chartier, 813 F.3d 891, 905 (9th Cir. 2016) (The movie "*The Hurt Locker* is not speech proposing a commercial transaction. Accordingly, our precedents relying on the lesser protection afforded to commercial speech are inapposite."); De Havilland v. FX Networks, LLC, 21 Cal. App. 5th 845, 850 (Cal. Ct. App. 2018) (The fact that "creative works generate income for their creators does not diminish their constitutional protection."); Kirby v. Sega of Am., Inc., 144 Cal. App. 4th 47, 58 (2006) (The protections of the First Amendment extend to "music, films, paintings, and entertainment, whether or not sold for a profit.").

8. *See*, *e.g.*, CAL. CIV. CODE § 3344.1(a)(1)(B)(i) (2024); OHIO REV. CODE ANN. § 2741.09 (1999); ARK. CODE ANN. § 4-75-1110 (2016); N.Y. CIV. RIGHTS LAW § 50-f(2)(d)(i) (2020); LA. STAT. ANN. § 51:470.5 (2022).

vast swaths of speech, including art, humor, political commentary, journalism, and criticism, all of which are the lifeblood of a free and well-functioning democracy, allowing us to debate, scrutinize, and laugh at the world around us. And absent such exemptions, right of publicity defendants sued over references to and depictions of real people would bear the burden of asserting their First Amendment rights as affirmative defenses in individual cases, an expensive and arduous task that itself chills speech.

Despite the inapplicability of right of publicity laws to uses in expressive works, individuals unhappy with their portrayals in such works nonetheless sometimes assert such claims (typically invoking statutes or common-law doctrines that lack explicit expressive-works exemptions). The courts just as routinely hold that the First Amendment bars these attempts at censorship, though often only after lengthy and expensive litigation.[9]

Several new forms of harm stemming from unauthorized uses of an individual's image or likeness have led to calls for new legislation to address uses outside the advertising/merchandising context. First, so-called "deepfake pornography," or "non-consensual intimate images." As with other internet-related technologies, the pornography sector was among the early adopters.[10] In late 2017, videos in which the faces of prominent actresses and singers including Gal Gadot, Scarlett Johansson, Taylor Swift, and Aubrey Plaza were swapped onto the bodies of pornographic actresses in explicit scenes began proliferating on the internet.[11] Second, the replication of actors' or singers' likenesses or voices in new works in which they did not actually perform. While the potential for such a phenomenon had been recognized decades earlier,[12] and a short film that premiered in 1987 at the Canadian Engineering Centennial Convention actually featured "computer-generated and synthetic" versions

9. *See, e.g.*, Porco v. Lifetime Ent. Servs., LLC, 195 A.D.3d 1351 (N.Y. App. Div. 2021) (holding that the First Amendment barred claim by convicted murderer over portrayal in docudrama, but only after *eight years* of litigation in New York state courts, including multiple appeals and an initial court order—later overturned—that barred broadcast of television movie); *De Havilland*, 21 Cal. App. 5th at 845 (First Amendment barred claim by actress over portrayal in docudrama (reversing trial court order allowing claim to proceed)); *Sarver*, 813 F.3d at 896 (First Amendment barred claim by individual allegedly portrayed in movie *The Hurt Locker*); Tyne v. Time Warner Ent. Co., L.P., 901 So.2d 802 (Fla. 2005) (First Amendment barred claims involving movie *The Perfect Storm*); Matthews v. Wozencraft, 15 F.3d 432, 439 (5th Cir. 1994) ("Courts long ago recognized that a celebrity's right of publicity does not preclude others from incorporating a person's name, features, or biography in a literary work, motion picture, news or entertainment story. Only the use of an individual's identity in advertising infringes on the persona." (quoting George M. Armstrong, Jr., *The Reification of Celebrity: Persona as Property*, 51 LA. L. REV. 443, 467 (1991)).

10. *See, e.g.*, Michael Brooks, *The Porn Pioneers*, THE GUARDIAN (Sep. 29, 1999)', https://www.theguardian.com/technology/1999/sep/30/onlinesupplement [https://web.archive.org/web/20260207001545/https://www.theguardian.com/technology/1999/sep/30/onlinesupplement] (describing pornography industry's early adoption of internet technologies).

11. *See* Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked*, VICE (Dec. 11, 2017), https://www.vice.com/en/article/gal-gadot-fake-ai-porn/ .[https://web.archive.org/web/20260207001752/https://www.vice.com/en/article/gal-gadot-fake-ai-porn/].

12. *See* Joseph J. Beard, *Casting Call at Forest Lawn: The Digital Resurrection of Deceased Entertainers—A 21st Century Challenge for Intellectual Property Law,* 8 HIGH TECH. L.J. 101, 102 (1993).

of Marilyn Monroe and Humphrey Bogart,[13] it was the rapid developments in generative artificial technologies beginning in the early 2020s that fueled alarm among actors and recording artists, as "deepfake Tom Cruise"[14] and a song sung by "Fake Drake" and a voice clone of The Weeknd[15] went viral on the internet.[16] And third, politicians raised concerns that digital replicas of them doing something they never did, or saying something they never said, could be deployed by political opponents, or pranksters, to harm their election prospects.[17]

Why were existing state right of publicity and related laws inadequate to address such bad acts? At least two reasons. First, many state right of publicity laws are by their terms limited to commercial uses, either through the scope of the right itself,[18] or via "expressive works exemptions" that explicitly carve out uses of one's likeness in movies, television programs, books, songs, plays, etc. from the ambit of such statutes.[19] Second, even where state right of publicity statutes were not explicitly limited to commercial uses, as a practical matter, enforcement actions to remove material containing unlawful digital replicas or deepfakes could not be undertaken against platforms (YouTube, Twitter/X, Instagram, Facebook, etc.) that hosted them, due to the protections afforded to such platforms by section 230 of the Communications Decency Act of 1996.[20] While section 230 contains a carve-out for "any law pertaining to intellectual property,"[21] some courts—including the U.S. Court of Appeals for the Ninth Circuit, which includes California, where many performers live, and where most of the major social

13. *See id.* at 104 (citing Nadia Magnenat-Thalmann & Daniel Thalmann, *The Direction of Synthetic Actors in the Film* Rendevouz à Montréal, IEEE COMPUT. GRAPHICS & APPLICATIONS, Dec. 1987, at 9).

14. *See* Rachel Metz, *How a Deepfake Tom Cruise on TikTok Turned into a Very Real AI Company*, CNN BUS. (Aug. 6, 2021), https://www.cnn.com/2021/08/06/tech/tom-cruise-deepfake-tiktok-company [https://web.archive.org/web/20260208225121/https://www.cnn.com/2021/08/06/tech/tom-cruise-deepfake-tiktok-company].

15. *See* Joe Coscarelli, *An A.I. Hit of Fake "Drake" and "The Weeknd" Rattles the Music World*, N.Y. TIMES (April 19, 2023), https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html [http://web.archive.org/web/20230419210519/https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html/].

16. *See Who Owns You? SAG-AFTRA Steps Up the Fight to Ensure Members Have Control of Their Own Likenesses,* SAG-AFTRA MAG., Summer 2018, at 30–33, https://digital.copcomm.com/i/1012073-summer-2018/31?m4= [https://perma.cc/MY7X-UC73?type=image].

17. *See, e.g.*, Shannon Bond, *How AI Deepfakes Polluted Elections in 2024,* NAT'L PUB. RADIO (Dec. 21, 2024), https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections [https://web.archive.org/web/20260207003848/https://www.npr.org/2024/12/21/nx-s1-5220301/deepfakes-memes-artificial-intelligence-elections] (reporting on phone calls to New Hampshire voters using fake "Joe Biden" voice and urging them not to vote in primary election).

18. *See, e.g.*, ARK. CODE § 4-75-1108 ("[A] person who commercially uses the name, voice, signature, photograph, or likeness of an individual is liable to the holder of the property right provided by this subchapter . . ."); N.Y. CIV. RIGHTS LAW § 51 (limited to uses "for advertising purposes or for the purposes of trade").

19. *See, e.g.*, CAL. CIV. CODE § 3344.1(a)(1)(B)(i) (2024); OHIO REV. CODE ANN. § 2741.09 (1999); ARK. CODE ANN. § 4-75-1110 (2016); N.Y. CIV. RIGHTS LAW § 50-f(2)(d) (2020); LA. STAT. ANN. § 51:470.5 (2022).

20. 47 U.S.C. § 230(c)(1) ("No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.").

21. 47 U.S.C. § 230(e)(2).

media platforms are headquartered and often mandate venue for disputes[22]—have interpreted that carve-out to apply only to *federal* intellectual property claims, meaning that the platforms face no liability for state right of publicity claims based on user-posted material, and thus are under no legal obligation to remove material that violates such state laws.[23]

II. NEW LEGISLATION

The last decade has seen the introduction, and in many cases enactment, of legislation intended to address the three distinct forms of harm noted above.[24] While the statutes vary in scope, as of September 2025 at least forty-five states had enacted laws addressing deepfake pornography.[25] And in May 2025, President Trump signed into law the "TAKE IT DOWN Act," federal legislation that criminalizes the nonconsensual publication of intimate images, including both actual photographs[26] as well as "digital forgeries" (i.e., deepfakes), and requires platforms to remove such material within 48 hours of receipt of a notice.[27] Regarding election-related deepfakes, "[b]y the end of 2024, twenty states had election deepfake laws, two of them outlawing such media material even if it was labeled not authentic."[28]

New York was the pioneer in enacting legislation seeking to protect actors and recording artists from unauthorized replication of their likeness and voices in new performances. By enacting Civil Rights Law § 50-f in 2020, the New York Legislature created a new postmortem right of publicity that covered commercial uses of a "deceased personality's" name, voice, signature, photograph, or likeness, for advertising or merchandising purposes, for forty years after the personality's death.[29] But separately, this new statute contained in § 50-f(2)(b) a novel right against deceptive uses

22. *See, e.g.*, *Terms of Service*, YOUTUBE (Dec. 15, 2023), https://www.youtube.com/t/terms [https://web.archive.org/web/20260221144433/https://www.youtube.com/t/terms] (designating the "federal or state courts of Santa Clara County, California" as forum for disputes).

23. *Compare* Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102 (9th Cir. 2007) (IP carve-out in § 230(e)(2) applies only to federal IP claims), *with* Hepp v. Facebook, 14 F.4th 204, 210 (3d Cir. 2021) (Section 230's IP carve-out applies to both federal and state IP claims.).

24. *See supra* Part I.

25. *Tracker: State Legislation on Intimate Deepfakes*, PUB. CITIZEN (updated Oct. 20, 2025), https://www.citizen.org/article/tracker-intimate-deepfakes-state-legislation/ [https://web.archive.org/web/20260221174741/https://www.citizen.org/article/tracker-intimate-deepfakes-state-legislation/].

26. The non-consensual dissemination of actual photographs of an individual who is nude or engaging in sexual conduct is often referred to as "revenge porn."

27. *See* Pub. Law No. 119-12, 139 Stat. 55, 119th Congress (May 19, 2025); *see also* Victoria Killion, *The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images*, CONG. RSCH. SERV. (May 20, 2025), https://www.congress.gov/crs-product/LSB11314 [https://web.archive.org/web/20260211145324/https://www.congress.gov/crs-product/LSB11314].

28. Chris Hables Gray, *Political Deepfakes and Elections*, FREE SPEECH CTR., MIDDLE TENN. STATE UNIV. (updated Jan. 11, 2025), https://firstamendment.mtsu.edu/article/political-deepfakes-and-elections/# [https://web.archive.org/web/20260221175615/https://firstamendment.mtsu.edu/article/political-deepfakes-and-elections/].

29. N.Y CIV. RIGHTS LAW § 50-f(2)(a), (8).

of a digital replica of a deceased performer in expressive works, where the performer died domiciled in New York. The new right created by the 2020 law was narrow and is more accurately described as a consumer-fraud provision than an intellectual property right. The provision applied only to uses of a "digital replica" (as defined in § 50-f(1)(c)) in a "scripted audiovisual work as a fictional character or for the live performance of a musical work."[30] And it only applied "if the use is likely to deceive the public into thinking it was authorized by" specified heirs of the deceased performer.[31] (As described below, the 2020 law was significantly amended in 2025.)[32]

Louisiana followed suit in 2022 with the Allen Toussaint Legacy Act, named after the New Orleans jazz musician whose relatives, following Toussaint's death, objected to vendors selling t-shirts and beer koozies depicting their departed loved one.[33] The Louisiana statute contains a traditional right of publicity (i.e., limited to commercial uses) that applies during an individual's life plus up to fifty years after. But it also includes a provision, which applies only to living individuals, barring "use [of] a digital replica in a public performance of a scripted audiovisual work, or in a live performance of a dramatic work, if the use is intended to create, and creates, the clear impression that the professional performer is actually performing in the role of a fictional character."[34]

The floodgates opened in 2024, with broad digital replica legislation introduced in at least ten states, and enacted in three: California, Tennessee, and Illinois. California's A.B. 1836[35] tore down the firewall in that state's post-mortem right of publicity statute, Civil Code section 3344.1, that had unambiguously excluded expressive works from the scope of the statute.[36] As amended by A.B. 1836, section 3344.1 now provides a cause of action against one "who produces, distributes, or makes available the digital replica of a deceased personality's voice or likeness *in an expressive audiovisual work or sound recording* without prior consent from the relevant right holder," albeit with exceptions for uses typically protected by the First Amendment, including comment, criticism, scholarship, satire, parody, and depictions in biopics and docudramas.[37] Tennessee's ELVIS Act ("Ensuring Likeness Voice and Image Security Act"), enacted following a heavy lobbying effort by that state's influential music industry, greatly expanded its existing right of publicity statute, adding "voice" to its scope of protection and creating a broad new cause of action against one who "publishes, performs, distributes, transmits, or otherwise makes available to the public an individual's voice or likeness,

30. *Id.* at § 50-f(2)(b) (with definitions at § 50-f(1)(c)).

31. *Id.*

32. *See infra* note 44 and accompanying text.

33. *See Meeting of Apr. 30, 2019, H. Comm. on Civ. Law & Proc.*, 2019 Reg. Sess. (La. 2019) (statement of Tim Kappel, Louisiana Music Creators, in support of H.B. 377 (Allen Toussaint Legacy Act)).

34. LA. STAT. ANN. §§ 51:470.1–470.6 (2022), also known as the Allen Toussaint Legacy Act.

35. A.B. 1836, 2023–24 Sess. (Cal. 2024).

36. *See* CAL. CIV. CODE § 3344.1(a)(1)(B) (exempting uses in "a play, book, magazine, newspaper, musical composition, audiovisual work, radio or television program, single and original work of art, work of political or newsworthy value, or an advertisement or commercial announcement for any of these works . . . if it is fictional or nonfictional entertainment, or a dramatic, literary, or musical work.").

37. CAL. CIV. CODE § 3344.1(a)(2)(A)(i)–(ii) (emphasis added).

with knowledge that use of the voice or likeness was not authorized by the individual."[38] Tennessee's right lasts for the life of the individual and extends postmortem for a term of potentially *forever*, assuming continuing exploitation of the right.[39] While the statute includes a set of exemptions intended to protect free expression,[40] the ELVIS Act considerably muddied the waters by providing that those exemptions now apply only "[t]o the extent such use [of one's NILV] is protected by the First Amendment to the United States Constitution."[41] This new language would essentially require a full First Amendment analysis to determine whether the statutory exemptions apply, which undermines the very purpose of the exemptions: to provide clarity so that parties can easily determine whether the statute applies to proposed conduct. And Illinois enacted H.B. 4875, which amended its existing right of publicity law to cover uses "for the purposes of distributing, transmitting, or otherwise making available a sound recording or audiovisual work that contains a simulated or artificially created version of an individual's identity, as a substitute for, in place of, or in a competitive fashion with, a sound recording or audiovisual work the individual would otherwise have personally created using the individual's identity."[42]

In 2025, Montana enacted its own digital replica statute, modeled on Tennessee's ELVIS Act, though with a shorter (twenty-year) post-mortem term and application only to Montana domiciliaries.[43] New York's legislature also enacted a bill amending its 2020 post-mortem right of publicity statute to create a true intellectual property right that applies to expressive works rather than an anti-fraud provision.[44] Also, California, Illinois, and New York have enacted bills that do not themselves create new digital replica rights, but instead establish rules governing the licensing of such rights; a similar bill is pending in Massachusetts.[45] In short, these bills require that a digital replica license include a "reasonably specific description of the intended uses of the digital replica," unless the individual is represented by counsel or by a union whose "collective bargaining agreement expressly addresses uses of digital replicas."[46] (The

38. TENN. CODE ANN. § 47-25-1105(a)(2); *see* Press Release, Office of the Governor, Tennessee First in the Nation to Address AI Impact on Music Industry (Jan. 10, 2024), https://www.tn.gov/governor/news/2024/1/10/tennessee-first-in-the-nation-to-address-ai-impact-on-music-industry.html [https://web.archive.org/web/20260324234619/https://www.tn.gov/governor/news/2024/1/10/tennessee-first-in-the-nation-to-address-ai-impact-on-music-industry.html].

39. TENN. CODE ANN. § 47-25-1104.

40. TENN. CODE ANN. § 47-25-1107.

41. *Id.* at (a).

42. H.B. 4875, 103rd Gen. Assem., Reg. Sess. (Ill. 2023–24).

43. H.B. 513, 69th Leg. (Mont. 2025).

44. 2025-S8391, 2025–26 Reg. Sess. (N.Y. 2025).

45. *See* A.B. 2602, 2023–24 Sess. (Cal. 2024); H.B. 4762, 103d Gen. Assemb., Reg. Sess. (Ill. 2024), amended by H.B. 3178, 104th Gen. Assemb., Reg. Sess. (Ill. 2025); S.B. 7676-B, 2023–24 Gen. Assemb. (N.Y. 2024); H.B 74, 194th Gen. Ct., Reg. Sess. (Mass. 2025); *see also* H.B. 388, 2025–26 Gen. Assem., Reg. Sess. (Vt. 2025) (a similar bill introduced in Vermont which failed to advance).

46. *See e.g.*, CAL. LAB. CODE § 927(a)(2)(A)(3) (enacted via A.B. 2602 (2024)).

topic of digital replicas is indeed addressed in the 2023 collective bargaining agreement between producers and SAG-AFTRA, the union representing actors.)[47]

Also in the 2024–2025 period, at least twelve states introduced but, at least so far, failed to enact, broad right of publicity/digital replica bills that would apply to uses in expressive works, including in Texas,[48] Mississippi,[49] Kentucky,[50] Louisiana,[51] New Jersey,[52] New Mexico,[53] Georgia,[54] Maryland,[55] Oklahoma,[56] South Carolina,[57] Ohio,[58] and Massachusetts.[59]

At the federal level, the NO FAKES Act, a bill that would establish a new federal intellectual property right in one's voice and likeness, including against uses in expressive works, was introduced by a bipartisan group of senators in 2024 and re-introduced with some changes in 2025.[60] The rights under NO FAKES would last for the life of the depicted individual plus seventy years (mirroring the term of copyright, but with earlier termination in the event of a specified period of nonuse) and includes exceptions for uses including parody, satire, commentary, and to depict the "individual as the individual in a documentary or in a historical or biographical manner, including some degree of fictionalization," though with a carve-out for deceptive uses.[61] And it includes a comprehensive intermediary-liability regime, roughly modeled on section 512 of the Digital Millennium Copyright Act, providing a notice-and-takedown process and a safe harbor shielding from liability online services that promptly disable access to works containing unauthorized digital replicas upon receipt of notices.[62] The NO FAKES Act has been endorsed by a wide array of stakeholders with an interest in regulation of digital replicas, including the Recording Industry Association of America (representing record labels), the Recording Academy (recording artists), SAG-AFTRA, the Motion Picture Association (major motion picture and television studios), OpenAI,

47. *See Digital Replicas 101, SAG-AFTRA* (2024), https://www.sagaftra.org/sites/default/files/sa_documents/DigitalReplicas.pdf [https://web.archive.org/web/20250406043659/https://www.sagaftra.org/sites/default/files/sa_documents/DigitalReplicas.pdf].

48. H.B. 3950, 89th Leg. (Tex. 2025).

49. S.B. 2778, 2024 Reg. Sess., Miss. Gen. Assemb. (Miss. 2024).

50. S.B. 317, 2024 Reg. Sess., Ky. Gen. Assemb. (Ky. 2024).

51. S.B. 6, 2024 Reg. Sess., La. Gen. Assemb. (La. 2024) (passed after amended to cover only sexually explicit images).

52. A. 4480, 2024–25 Reg. Sess. (N.J. 2024).

53. H.B. 221, 57th Leg., 2025 Reg. Sess. (N.M. 2025).

54. H.B. 566, 159th Gen. Assemb., Reg. Sess. (Ga. 2025).

55. H.B. 1407, 2025 Reg. Sess., Md. Gen. Assemb. (Md. 2025).

56. H.B. 2216, 2025 Reg. Sess., Okla. Gen. Assemb. (Okla. 2025).

57. H.B. 3404, 126th Gen. Assemb., 2025–26 Reg. Sess. (S.C. 2025).

58. H.B. 185, 136th Gen. Assem., Reg. Sess. (Ohio 2025).

59. H.B. 1751, 194th Gen. Ct., Reg. Sess. (Mass. 2025).

60. Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2025 ("NO FAKES Act"), S. 1367, 119th Cong. (2025).

61. *Id.*

62. 17 U.S.C. § 512.

the National Association of Broadcasters, Authors Guild, the National Center on Sexual Exploitation, Television Academy, Google/YouTube, IBM, and others.[63]

III. LOOKING FORWARD

In 2019, the law around depiction of individuals in motion pictures and television programs appeared settled. Cases like *Sarver*, *De Havilland*, and *Tyne* held clearly and emphatically that right of publicity law could not be used by individuals to censor portrayals of them, whether because they did not like those portrayals or they simply wanted to be paid for them.[64] But no longer. New "digital replica" laws enacted in California, New York, Tennessee, Illinois, and Montana, and proposed in Congress and more than a dozen other states, have upended once settled law, creating significant uncertainty around the ability to use new technology, fueled by advances in artificial intelligence, to depict people on screen. While these laws or bills typically include exemptions intended to protect First Amendment rights, those exemptions' clarity is blurred by vague language and carve-outs, and their application has yet to be tested by the courts.[65]

The lengthy post-mortem terms of such laws and bills are especially concerning, as they could affect the ability of filmmakers to tell stories about historical figures, particularly deceased performers, many of whose rights are aggressively exploited and protected by corporate rights holders, as well as other deceased individuals.[66] And they are of questionable constitutionality. As with any law regulating the content of speech,[67] right of publicity laws are subject to strict First Amendment scrutiny, which

63. *Blackburn, Coons, Salazar, Dean, Colleagues Introduce "NO FAKES Act" to Protect Individuals and Creators from Digital Replicas*, OFF. OF SEN. MARSHA BLACKBURN (Apr. 9, 2025), https://www.blackburn.senate.gov/2025/4/technology/blackburn-coons-salazar-dean-colleagues-introduce-no-fakes-act-to-protect-individuals-and-creators-from-digital-replicas [https://web.archive.org/web/20260127233119/https://www.blackburn.senate.gov/2025/4/technology/blackburn-coons-salazar-dean-colleagues-introduce-no-fakes-act-to-protect-individuals-and-creators-from-digital-replicas].

64. De Havilland v. FX Networks, LLC, 21 Cal. App. 5th 845, 871 (Cal. Ct. App. 2018); Sarver v. Chartier, 813 F.3d 891, 891–92 (9th Cir. 2016); Tyne v. Time Warner Ent. Co., 901 So.2d 802, 802 (Fla. 2005).

65. As of this writing, the author is unaware of any lawsuits filed under the digital replica provisions of the five new such laws enacted since 2020.

66. CMG Worldwide, a leader in this sector, lists approximately 270 "clients," including deceased actors James Dean, Bette Davis, Ingrid Bergman, Jimmy Stewart, and John Wayne; late sports figures from Arthur Ashe to Andre the Giant to Lou Gehrig; historical figures including Albert Einstein, Amelia Earhart, Frank Lloyd Wright, General George S. Patton Jr., Black Panther Huey P. Newton, Oscar Wilde, Rosa Parks, and Thomas Edison. *See Client List,* CMG WORLDWIDE (Aug. 2025), https://www.cmgworldwide.com/wp-content/uploads/2025/08/CMG-Worldwide-Client-List-25.pdf [https://web.archive.org/web/20250913072612/https://www.cmgworldwide.com/wp-content/uploads/2025/08/CMG-Worldwide-Client-List-25.pdf].

67. *Sarver,* 813 F.3d at 903 ("By its terms, California's right of publicity law clearly restricts speech based upon its content."); *see also* Eugene Volokh, *Freedom of Speech and the Right of Publicity*, 40 HOU. L. REV. 903, 912 n.35 (2003) ("The right of publicity is clearly content-based: It prohibits the unlicensed use of particular content (people's name or likenesses) . . . But even if it's seen as content-neutral, strict scrutiny is still the proper test, because the right of publicity doesn't leave open ample alternative channels for the speaker to convey the content that he wishes to convey.").

means that they are "presumptively unconstitutional" and can survive constitutional challenge only upon a showing that they: 1) serve a compelling government interest; and 2) are narrowly tailored to serve that interest.[68] A court could determine that certain unconsented uses of digital replicas to replace *living* actors or recording artists could interfere with their ability to earn a living, establishing a compelling state interest sufficient to satisfy the constitutional requirement. However, that employment-based interest does not exist for deceased individuals. And other purported justifications for protecting deceased performers are unavailing. Any interest in a performer's reputation or dignity is already governed by defamation and privacy law, which is personal to the individual at issue.[69] But recognizing dignitary interests of deceased individuals, and giving heirs or corporate successors the ability to sue over them, would represent a radical change in centuries of American law, under which "there can be no defamation of the dead."[70]

IV. CONCLUSION

Right of publicity law is in the beginning stages of a revolution, reversing the last several decades' trend of limiting its application, via statutory expressive-works exemptions as well as case law, to commercial (i.e., advertising/merchandising) uses. A handful of states have already enacted broad new laws, and Congress is actively debating one with national application. Litigation testing the scope, and constitutionality, of these new laws will inevitably follow. And it may take another several decades for the law governing the use of digital replicas to settle into a new equilibrium.

68. Reed v. Town of Gilbert, Ariz., 576 U.S. 155, 163 (2015).

69. *See* Gertz v. Robert Welch, Inc., 418 U.S. 323, 341 (1974) ("The legitimate state interest underlying the law of libel is the compensation of individuals for the harm inflicted *on them* by defamatory falsehood.") (emphasis added); Lugosi v. Universal Pictures, 25 Cal. 3d 813, 821 (1979) ("It is well settled that the right of privacy is purely a personal one; it cannot be asserted by anyone other than the person whose privacy has been invaded, that is, plaintiff must plead and prove that *his* privacy has been invaded.").

70. Restatement (Second) of Torts § 560, comment a (1977); *see also*, *e.g.*, Bradt v. New Nonpareil Co., 108 Iowa 449 (1899) ("The rule that an heir may recover for a libel of one deceased does not seem to have gained a foothold in this country, and we know of no principle that will sustain such an action."); Meeropol v. Nizer, 381 F. Supp. 29, 34, 35 n.3, 37 (S.D.N.Y. 1974), *aff'd*, 560 F.2d 1061 (2d Cir. 1977) (rejecting defamation and invasion of privacy claims by children of convicted and executed spies Julius and Ethel Rosenberg over statements in book because such claims expire upon the death of the subject of the statements at issue); Kelly v. Johnson Publishing Co., 160 Cal. App. 2d 718, 723 (1958) ("Defamation of a deceased person does not give rise to a civil right of action at common law in favor of the surviving spouse, family, or relatives, who are not themselves defamed. A libel on the memory of a deceased person is not deemed to inflict on the surviving relatives of the deceased any such legal damage as will sustain a civil action for the defamation.").

Deepfakes and Private Rights in the Perspective of EU Law: Is It Necessary to Intervene?

Valérie-Laure Benabou*

* Valérie Laure Benabou is Professor at the University of Versailles-Saint Quentin, Paris-Saclay. A Member of the D@NTE Research Center, she is also a Member of the Superior Council of Literary and Artistic Property within the French Ministry of Culture.

INTRODUCTION

Fakes have always existed, but AI systems' now-limitless capacity for manipulation has also uncovered positive uses: Our capacity to now reconstitute events or places that once existed but were never recorded is only an illustration of what can also be an interesting development of these new fakes. The rapid spread of AI-generated images or voices that starkly—deeply—resemble the individuals on which they have been modeled has raised two major, and quite opposing, concerns. On the one hand is the risk of deceiving the public: giving them the false belief that what is been displayed is an expression of reality. On the other hand is the development and control of the emerging market for these new synthetic features. The first concern ought to produce protective measures to prevent the risk of deception, but such measures run the risk of impeding the potential development of the market at a moment where the still-

unrealized potential of the technology provides both new opportunities for creating "fakes" and genuine benefits.

I. THE LACK OF SPECIFIC HARMONIZED SOLUTIONS FOR DEEPFAKES

Part I will address the (so far limited) EU obligation of Disclosure imposed on AI deployers by the AI Act (I.A), recall that due to the restricted competence of the European Union, there is a lack of harmonization of the criminal offenses that can cover some misuses of deepfakes (I.B), and examine a recent political initiative driven by the Danish government aiming at creating an exclusive right on individual's likeness (I.C).

A. THE (SO-FAR LIMITED) EU OBLIGATION OF DISCLOSURE FOR AI DEPLOYERS

With the exception of rules regarding unfair business practices and consumer protection,[1]European law has so far addressed the risk of deception primarily by imposing a specific disclosure obligation on deployers (i.e., professional users) of deepfakes. Deployers must clearly and distinguishably indicate that the content has been artificially generated or manipulated, typically by labeling the AI output accordingly and disclosing its artificial origin.[2] The definition appears in Article 3, paragraph 60 of the AI Act, which provides that "deepfake means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful."[3] This legal intervention may be considered minimalist in some respects. First, it has not yet entered into force but will only apply from August 2026. Second, the definition is limited to images, audio and video content, leaving aside deception that would be expressed through literary form, such as fake news,[4] even though a similar

1. Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market, 2005 O.J. (L 149) 22, as amended by Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019, 2019 O.J. (L 328) 7 (regarding the better enforcement and modernization of EU consumer protection rules).

2. Artificial Intelligence Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 ["EU AI Act"], 2024 O.J. (L 1689), recital 134 and art. 50(4). "Deployers of an AI system that generates or manipulates image, audio or video content constituting a deepfake, shall disclose that the content has been artificially generated or manipulated. This obligation shall not apply where the use is authorised by law to detect, prevent, investigate or prosecute criminal offence. Where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations set out in this paragraph are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work."

3. EU AI Act, *supra* note 2, art. 3(60).

4. *Id.* art. 50(4). The second paragraph of art. 50(4) provides that "deployers of an AI system that generates or manipulates text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. This obligation shall

disclosure obligation exists for manipulated informational content. More fundamentally, the AI Act also does not prohibit deepfakes, nor does it establish a way of distinguishing truth from manipulation, which would be the only solution adequately addressing the risk of deception normally at the heart of this disclosure obligation. This obligation is subject to limitations: It does not even apply when such use is authorized by law to detect, prevent, investigate, or prosecute criminal offenses. It is also limited when the content forms part of an evidently artistic, creative, satirical, fictional, or analogous work or program, in which case the disclosure should be made in an appropriate manner that does not hamper the display or enjoyment of the work. This possible contextualization demonstrates that deepfakes are not treated as wrongful in themselves, notably when their use is embedded in a fictional setting: What matters, rather, is ensuring the public's awareness that it is being confronted with a synthetic output. As Magritte's *The Treachery of Images* reminds us, the representation of a pipe is not a pipe; likewise, the algorithmic representation of a thing or a person is never that thing or person.

B. THE LACK OF HARMONIZATION OF CRIMINAL OFFENSES

This form of moral neutrality demonstrates an absence of choice with regards to the direction that the European Union should take in regulating deepfakes. Although the AI Act does not proscribe deepfakes, prohibition of their use and creation can nevertheless rely on public-interest considerations, such as preventing forgeries committed through synthetic features for purposes of extorsion, the circumvention of authentication systems, revenge pornography, or similar harms. Generally, these behaviors are already banned irrespective of the presence of deepfakes, which in these cases would only be considered as a means of committing the offense.

Yet, in some cases, national legislation may directly ban deepfakes as such, as illustrated by French criminal law. Until recently, Article 226-8 of the penal code punished the act of "publishing, by any means whatsoever, the montage made with the words or image of a person without his consent, if it does not appear obvious that it is a montage or if it is not expressly mentioned," by one year's imprisonment and a fine of 15,000 euros.[5] Since the enactment of Loi SREN,[6] which has modified the article 226-8 in 2024, it is now also prohibited to "bring to the attention of the public or a third party, by any means whatsoever, visual or sound content generated by algorithmic processing and representing the image or words of a person, without his consent, if it is not obvious that it is algorithmically generated content or if it is not expressly

not apply where the use is authorized by law to detect, prevent, investigate or prosecute criminal offences or where the AI-generated content has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content."

5. CODE PÉNAL [Penal Code] art. 226-8 (Fr.) (before 2024 amendment).

6. Loi 2024-449 du 21 mai 2024 visant à sécuriser et à réguler l'espace numérique [To Secure and Regulate the Digital Space], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], May 21, 2024.

mentioned."[7] Recently, Italy has adopted new provisions in its Criminal Code, introducing Article 612-quater, entitled "[i]llicit dissemination of content generated or altered using artificial intelligence," according to which "[a]nyone who causes unjust harm to a person by transferring, publishing, or otherwise disseminating, without her consent, images, videos, or voices that have been falsified or altered through the use of artificial intelligence systems and are capable of misleading as to their authenticity, shall be punished with imprisonment for a term of one to five years."[8] In Italy, the offense is normally punishable upon complaint by the offended person.

Although the two provisions are broadly similar, their elements differ slightly: Where the definition of the French offense relies only on the absence of consent of the person targeted by the deepfake and on the absence of disclosure of the algorithmic origin of the feature, the Italian law requires the existence of unjust ham for the victim, acts of falsification or alteration and finally the capacity to deceive on the authenticity of the information conveyed by the deepfake. These solutions are not harmonized because they rely on criminal law, which does not fall within the scope of the competence of the European Union.

C. RECENT POLITICAL INITIATIVES

Facing this lack of uniformity within the EU, a call for action came recently from the Danish government, which took over the presidency of the EU in July 2025. Denmark is currently adopting a modification of its copyright law to extend the system of exclusive rights to every individual's likeness, following the rationale of the No Fakes Act currently being discussed in the United States.[9] In doing so, Denmark expects to increase the efficiency of the notice and take down procedure when fighting against the unconsented use of deepfakes, by benefiting from the reactivity of the platform when copyright rules are invoked. It has signified its intention to promote such a system at the EU level.

---

7. CODE PÉNAL [Penal Code] art. 226-8 (Fr.) (amended 2024).

8. Art. 612-quater, Illecita diffusione di contenuti generati o alterati con sistemi di intelligenza artificiale, L. 132/2025, 10 October 2025 (It.) ("Chiunque cagiona un danno ingiusto ad una persona, cedendo, pubblicando o altrimenti diffondendo, senza il suo consenso, immagini, video o voci falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuinità, è punito con la reclusione da uno a cinque anni. Il delitto è punibile a querela della persona offesa. Si procede tuttavia d'ufficio se il fatto è connesso con altro delitto per il quale si deve procedere d'ufficio ovvero se è commesso nei confronti di persona incapace, per età o per infermità, o di una pubblica autorità a causa delle funzioni esercitate.") ["The offense is punishable upon complaint by the offended person. However, proceedings shall be initiated ex officio if the act is connected with another crime for which proceedings must be initiated ex officio, or if it is committed against a person who is incapable due to age or infirmity, or against a public authority due to the functions exercised."].

9. *See* Elin Hofverberg, *Denmark: Political Parties Agree to Protect Danes Against Deepfakes*, LIBR. OF CONG. (Aug. 5, 2025), https://www.loc.gov/item/global-legal-monitor/2025-08-05/denmark-political-parties-agree-to-protect-danes-against-deepfakes/ [https://web.archive.org/web/20250812222931/https://www.loc.gov/item/global-legal-monitor/2025-08-05/denmark-political-parties-agree-to-protect-danes-against-deepfakes/].

This initiative brings forth at least two types of questions. On the one hand, it raises the philosophical debate on the opportunity to create an intellectual property right for the benefit of all citizens, notwithstanding the absence of any creative input. On the other hand, it highlights the relevance of an economic monopoly, which is a right to prohibit and to authorize, eventually against remuneration, with the corollary risk of commodification of the likeness and voice of the individuals. It thus appears necessary to assess whether other types of protection of the right of the individuals, such as the protection of personality rights or the protection of personal data, would be more adequate to tackle the issue of personal deepfakes.

This Article will explore the opportunity to use the concept of ownership to deal with the deepfakes issues starting in the next part by appraising the existing solutions of IP (Part II). Then, it will consider the different potential consequences of the extension of such model, which, after scrutiny, does not prove to be necessary (Part III). Finally, it will envisage whether other types of existing protections based on the protection of the person within the EU could be more suitable (Part IV).

II. CAN EXISTING IP SOLVE THE DEEPFAKES ISSUE?

Part II will explore the possibility of using the concept of ownership in relation to deepfakes, and in this respect, the potential applicability of the existing intellectual property rights (II.A), together with the adequacy of the internal balance of interests within the IP system (II.B).

A. A PARTIAL RELEVANCE OF EXISTING INTELLECTUAL PROPERTY RIGHTS

Under the broad definition of deepfakes adopted by the AI Act, not only the image or voice of a person, but also objects, places, entities, or events, may constitute the subject-matter of a deepfake.[10] This AI-generated reconstitution may therefore concern persons, things, and information alike. Unsurprisingly, such reconstitution may infringe an intellectual property right when it entails the reproduction and/or adaption of protected content. Copyright (II.A.1), related rights (II.A.2), the sui generis database right (II.A.3) and even industrial property rights (II.A.4) may, under certain conditions, be exercised against the publication of a deepfake that would infringe the scope of the right holder's exclusivity. Conversely, they may also be licensed to authorize such use.

1. Copyright

Take the example of a photograph that is copyright-protected and modified through the use of an AI system by including an element that was not present in the original image or by removing part of the scenery. This type of manipulation is part of the history of photography, so much so that the relation between truth and fiction in photography has long raised debates between those who defend photographic work as

10. *See supra* notes 2–4 and accompanying text.

work of the mind, and those who claim that a photo is merely the automatic reflection of reality. In this respect, AI does not radically alter the state of the play, albeit for the quality and economy of the algorithmic process. In such a case, the modification of the photograph can be qualified at the very least as an act of reproduction, if not also as an act of adaptation, and reproduction in principle requires the prior consent of the author. In the absence of such authorization, and if the reproduction is not covered by a specific legal exception, both the creation of the deepfake and its communication to the public will infringe the copyright holder's monopoly. European copyright law has already harmonized the exclusive rights of reproduction, communication to the public, and distribution, so that the answer will, in principle, be uniform across Member States.[11] In such a situation, the modification of the work will appear as an act of reproduction thereof, this broad conception of the reproduction right having been reaffirmed by the European Court of Justice in the *Eva-Maria Painer*[12] and the *Art & Allposters*[13] cases. This is copyright business as usual.

More difficult is the case in which the deepfake output does not reproduce an identifiable part of a pre-existing work, but was conceived "à la manière de," or in the style of, a specific author, and is evocative of its works or personality. Given the classic distinction between form and idea, copyright law normally rejects the mere protection of style when no reproduction of the form of a work can be showed. Yet, for AI-generated deepfakes, it is likely that the works of the author in question have been reproduced at one point in the process of feeding the application (training, retrieval-augmented generation (RAG), fine-tuning), so that the principal characteristic of the paintings, photos, videos, songs can later be retrieved. Consequently, even if the comparison of the output with a specific work shows little or no resemblance, the possibility of recognizing the author's style may nevertheless imply prior digital acts of reproductions, whereas the forger in the analogue world was merely reconstructing the work mentally. The problem here lies in identifying the reproduction of a specific work, which is ordinarily a condition for bringing suit.

There also remains, uncovered under EU law (except for the database and for the computer program), the right of adaptation, whose definition may vary across countries. Under a narrow understanding, adaptation corresponds to a situation where a work created in one form of expression—say, a literary work—is transformed into another form of expression, such as a movie or a video game. AI now offers unexpected possibilities in this respect: AI applications can produce music out of a book, or images from a piece of music. Even if the result may appear surprising from the point of view of the human expectations, this type of transformation may be understood as adaptation, even though no formally recognizable elements of the original work will be found in the synthetic output. In such a case, a claim based on reproduction rights

11. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [hereinafter "Infosoc Directive"], arts. 2–4, 2001 O.J. (L 167).

12. Case C-145/10, Eva-Maria Painer v. Standard VerlagsGmbH and Others, 2011 E.C.R. I-12533.

13. Case C-419/13, Art & Allposters Int'l BV v. Stichting Pictoright, ECLI:EU:C:2015:27 (Jan. 22 2015).

may fall short, due to the impossibility of identifying the original work in the output. Although this difficulty is not limited to deepfakes, the absence of harmonization of the adaptation right constitutes a barrier to a coherent European framework. Intervention on the part of the EU would be pertinent, as it would help right holders ascertain the scope of their rights in relation to AI-driven, transformative uses.

2. Related Rights

Even though the practice of deepfakes is not limited to transformation of the image or voice of famous individuals, the most salient examples concern them (performers in particular). Possible uses are wide-ranging: synthetizing the voice of a comedian to offer a service of audiobooks allegedly read by the artist; allowing an actor to appear in a film while he cannot be physically present during the shooting, or sparing the use of a stuntman for dangerous scenes; resurrecting deceased performers to appear on screen or produce a new song (see the Beatles).[14] These possibilities are a source of concern for performers who fear the risk of substitution of their effective presence by these digital duplicates. Besides, the subject matter of their exclusive rights as defined in the Directive 2006/115 is the "performance."[15] Consequently, where a deepfake mimics an actress playing a role that she never actually performed, it is debatable whether her exclusive right may be exercised, since there is no actual reproduction, communication to the public, or distribution of her performance. Among other reasons, these uncertainties help explain why some legal systems are considering the adoption of new provisions within the IP regime in order to secure broader performers' rights.

In our view, it is not necessary to create a whole new IP regime to solve this problem if we retain a large interpretation of the notion of performance. It is obvious that the use of the image or the voice of a performer in deepfake or duplicate does not implicate her ordinary, day-to-day likeness or voice. Who would make a fake Janis Joplin song merely by reproducing the sound of her speaking voice from interviews, and not from the songs she already sang? Who would train a deepfake application on photographs of Marylin Monroe buying a hot dog, without make-up and a hairdo? Most of the examples mentioned above refer to situations where the commercial attractiveness of the deepfake comes from the extraction of the substantial value of the performer's prior interpretations and not merely from her personal attributes. Therefore, even if the transformative use made by the AI-generated deepfake results in a work that the artist never actually performed, this result cannot be achieved without the prior reproduction and selection of the characteristics of the performer *while performing*. The economic value of the service derives from the extraction of the value of those aggregated performances.

---

14. *See* Miguel Perez, *How Producers Used AI to Finish the Beatles' "Last" Song, "Now and Then,"* NPR MUSIC (Nov. 2, 2023), https://www.npr.org/sections/world-cafe/2023/11/02/1208848690/the-beatles-last-song-now-and-then [https://perma.cc/8BP5-FM2G].

15. Directive 2006/115/EC, of the European Parliament and of the Council of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property, 2006 O.J. (L 376) 28.

The same reasoning may be extended to producers' rights. If a deepfake mimics a previous film or song, its reproduction is plausible. For example, there is certainly a reproduction of the phonogram where the voice of a singer singing a specific song is reproduced in a video featuring another character pretending to interpret it. A more difficult case would be one in which the voice of a performer is being synthetized to interpret new songs. As with copyright, it is likely that the training necessary for the development of the application required copies of previous fixations. But in *Pelham*, the ECJ held that article 2(c) of Directive 2001/29/EC

> must, in the light of the Charter of Fundamental Rights of the European Union, be interpreted as meaning that the phonogram producer's exclusive right under that provision to reproduce and distribute his or her phonogram allows him to prevent another person from taking a sound sample, even if very short, of his or her phonogram for the purposes of including that sample in another phonogram, unless that sample is included in the phonogram in a modified form unrecognizable to the ear.[16]

3. Makers of a Sui Generis Database Right

A more unconventional hypothesis lies in the possibility to apply the maker's sui generis database right whenever the AI system reproduces data extracted from a database to generate deepfakes.[17] A database compiled by archeologists documenting the excavations of a certain site may be used to reconstitute through AI an image of this site as it might have appeared before its destruction. As the quantified-self movement grows, real-time data on activities, reporting on pulse, blood pressure, weight, or snoring during the night of a person could be also analyzed as a database of one's personal data that may be used to build a very sophisticate avatar of this person by an AI system.

Such an intellectual property right may prove useful in combating the undue appropriation of data because, subject to certain conditions, the maker of a database has the right to prevent the extraction and/or re-utilization, whether quantitative or qualitative, of the whole or of a substantial part of that database.[18] Interestingly, the European Court of Justice has interpreted this notion of extraction broadly, so as to encompass mere intellectual extraction, even in the absence of material or digital reproduction. In the *Directmedia* case, the ECJ held that the acknowledgement, by the maker of an anthology of German poems, that he had had access to a former anthology before making his own selection of poems was sufficient to amount to quantitative

16. Case C-476/17, Pelham I, ECLI:EU:C:2019:624 (July 29, 2019).

17. According to article 1(2) of Council Directive 96/9/EC of 11 March 1996 on the legal protection of databases, 1996 O.J. (L 77) 20, a database shall mean "a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means."

18. Council Directive 96/9/EC of 11 March 1996 on the legal protection of databases, art. 7(1), 1996 O.J. (L 77) 20 ("Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.").

extraction, due to the number of the same occurrences appearing in both anthologies.[19] In such a case, there was no need to demonstrate a prior act of reproduction in the course of making the database, as it was obvious from the content itself that such an extraction had occurred, due to the substantial presence of elements in common. It was sufficient that the value of the first maker's selection had been extracted. Even if the significance of this case should not be overstated,[20] it offers a means of bypassing the requirement for proof of technical reproduction. To return to the example of an AI-generated reconstitution of an ancient city, the fact that the archaeologist's database constitutes the only existing source of information could suffice to establish the existence of an extraction and, consequently, of an infringement of the sui generis right if no authorization had been given.

The difficulty, however, lies in demonstrating that the maker made a substantial investment in the obtention, verification, or presentation of the contents of the database. The Court of Justice has interpreted this condition narrowly across several judgments delivered in 2004: Investments made in the creation of the information are not relevant if the maker cannot prove any investments in the production of data itself.[21] Under this interpretation, the cost of the excavation itself cannot be taken into account: Only the time and money necessary to build the database from the results of the excavation are relevant. This requirement would likely constitute an obstacle for an individual seeking to claim ownership over the body of data she produced, for example, when running. One cannot be regarded as making a database of their own vital, bodily information simply by running. Only a person demonstrating investment in the obtention, verification or presentation of such data—namely, the service which collected the data— would be in a position to claim such ownership. Due to this requirement, using the sui generis right to combat the undue and massive appropriation of personal likeness seems unlikely.

4. Industrial Property Rights

It is also possible to claim ownership over a design or model or a trademark if the protected subject matter has been reproduced in the deepfake. Here again, the condition of identification of the element protected within the deepfake will be required. A video appearing to be a fake commercial for a famous trademark of soda would potentially be considered as an infringement if the use of the trademark has not been authorized by the right holder, subject to internal or external limitations for

19. C-304/07, Directmedia Publ'g GmbH v. Albert-Ludwigs-Universität Freiburg, 2008 E.C.R. I-7565, ¶ 60.

20. *Id.* at ¶ 14. In this case, the defendant recognized that he had access to this prior anthology while making his own one.

21. Fixtures Mktg. Ltd v. Oy Veikkaus Ab, Case C-46/02, 2004 E.C.R. I-10365 (Grand Chamber Nov. 9, 2004); British Horseracing Bd. Ltd. V. William Hill Org. Ltd., Case C-203/02, 2004 E.C.R. I-10415 (Grand Chamber Nov. 9, 2004); Fixtures Mktg. Ltd v. OPAP, Case C-444/02, 2004 E.C.R. I-10549 (Grand Chamber Nov. 9, 2004); Fixtures Mktg. Ltd v. Svenska Spel AB, Case C-338/02, 2004 E.C.R. I-10497 (Grand Chamber Nov. 9, 2004); *see also* Estelle Derclaye, *The European Court of Justice Interprets the Database Sui Generis Right for the First Time*. 30 EUR. L. REV. 420 (2005).

parody. Specifically, in the field of trademark law, the guarantee of origin—which is a fundamental function of the trademark recognized as such by the ECJ[22]—would ground the claim of the right holder whenever the use of the trademark is likely to cause a risk of confusion for the consumer. Even if the likelihood of confusion is not a legal condition applicable in the realm of protection of models, case-law could sometimes rely on such appraisal while analyzing the overall impression provided by the model. The recent extension of the definition of design and model to non-physical products now expands the protection to movement, transition, and animation that will certainly pave the way to more claims on that ground against deepfakes.[23]

In a nutshell, actual property rights may be applicable to deepfakes whenever they reproduce significant parts of the protected subject matter, which may be the case in various situations. IP is, therefore and in principle, relevant to tackle certain issues from the perspective of the right holder.

B. THE BALANCE OF THE INTERESTS WITHIN THE IP SYSTEM

The IP system cannot be reduced to a mere granting of exclusive rights but constitute a comprehensive mechanism aimed at combining different and sometimes divergent interests. Deepfakes may be made and are effectually used to convey a message, which, when it is not contrary to public order rules, may benefit from the protection of the freedom of expression. It is therefore interesting to assess whether the IP system contains appropriate rules to balance the respective interests. As many of these messages are humorous or critical, the application of the parody exception widely recognized in EU Intellectual property law may apply *de lege lata* (B.1). But since the public deception may increase due to AI sophisticated tools, the willingness of the right holder not to be associated a deceptive message may be taken into account in a reinforced way, which could be achieved, *de lege ferranda*, by an EU-wide recognition of the moral right (B.2).

1. IP and Parody

While safeguarding the truth is not a relevant justification for IP, and consequently not a duty imposed on the right holder, the protection of freedom of expression may, to some extent, limit the exercise of the exclusive right. This aspect is particularly important for deepfakes, some of which are created with the intention to mock or to criticize. Notions of parody, caricature, or pastiche are being developed within the IP system for copyright, related rights, and, lately, for designs and models with the recent

22. Terrapin (Overseas) Ltd. v. Terranova Industrie CA Kapferer & Co., Case C-119/75, 1976 E.C.R. 1039 (June 22, 1976); SA CNL-Sucal NV v. Hag GF AG (Hag II), Case C-10/89, 1990 E.C.R. I-3711 (Oct. 17, 1990).

23. Regulation (EU) 2024/2822 of the European Parliament and of the Council of 23 October 2024 amending Council Regulation (EC) No 6/2002 on Community designs and repealing Commission Regulation (EC) No 2246/2002, art. 3(1)–(2), 2024 O.J. (L 2822).

adoption of the EU package.[24] Such an extension has not yet been made regarding trademarks, where the satirical use of protected signs can nevertheless be tolerated through the application of the external principles of freedom of expression and/or freedom of creation. So far, the notion of parody has been interpreted in the realm of copyright by the ECJ in the famous *Deckmyn* case, where the Court decided that it should be an autonomous notion of EU law, thus discarding the competence of the Member States on this matter. [25]

Commenting on this decision, the advocate general in his last conclusions under the case *Pelham II* considered that this

> fairly broad approach to 'parody' offers some room for creative reuse of protected material. It may apply, for instance, to many instances of memes, involving the 'recognizable' reproduction of frames of films and their humoristic subversion through substantial modifications and/or addition of captions. Similarly, it may cover certain cases of mashups, potentially created through the 'sampling' of other phonograms, characterized by their humoristic tone or incongruity. It can also cover certain 'found footage', where frames of films are reused in a humoristic way, or even creative *détournement* of some beloved copyright-protected characters.[26]

No doubt that deepfakes may correspond to some elements of this non-exhaustive list.

The *Deckmyn* case has established the cumulative criteria required for applying the parody exception: The parody shall evoke the pre-existing work, albeit with noticeable differences, but must also show humor or mockery.[27] If the parody is made with a work, it is not necessarily about it. Paradoxically, the Court also held that the exception does not require the mention of the source of the work subject to the critical or ironic use.[28] Whereas the parody lies precisely in the differences between the source and the result, the judge did not consider important to let the public appraise these differences and to impose to the person responsible for the parody a duty to inform it. Even if parody exception may justify the rejection of the exclusive right, it does not facilitate the public knowledge of the "truth."

Interestingly, in the abovementioned conclusions in *Pelham II*, the advocate general also suggested a definition for the pastiche described as "an artistic creation that (i) evokes an existing work, adopting its distinctive "aesthetic language", while (ii) exhibiting perceptible differences from the source imitated, and (iii) is intended to be recognized as an imitation."[29] If certain deepfakes were to be recognized pastiche in respect of these criteria, the tribute to the work at the origin of the pastiche should be somehow communicated to the public. It should be indicated (in one way or another)

---

24. Directive (EU) 2024/2823 of the European Parliament and of the Council of 23 October 2024 on the legal protection of designs (recast), 2024 O.J. (L 2823); Regulation (EU) 2024/2822 of the European Parliament and of the Council of 23 October 2024 amending Council Regulation (EC) No 6/2002 on Community designs and repealing Commission Regulation (EC) No 2246/200, 2024 O.J. (L 2822).

25. Case C-201/13, ECJ, Johan Deckmyn v. Helena Vandersteen, ECLI:EU:C:2014:2132 (Sept. 3, 2014).

26. Opinion of AG Emiliou, *Pelham II*, Case C-590/23 at ¶ 94, EU:C:2025 (17 June 2025).

27. *Deckmyn*, Case C-201/13 (2014).

28. *Id.*

29. Opinion of AG Emilou, *Pelham II*, Case C-590/23 at ¶ 81.

or, at least, be recognizable as such by the viewer or listener who already knows the source. Consequently, they would be able to appreciate the distance between the deepfake and its source, and to clearly identify the nature of the pastiche. Thus, expressly or implicitly informed of the intent of the maker of the pastiche, the public would avoid any deception. Such a disclosure would be even more useful for the knowledge of the public than the requirements imposed by article 50 of the AI Act.[30]

Yet, with such a broad definition, pastiche would systematically limit the exclusive right of the right holder once the deployer of the deepfake claims his intention to make such a pastiche. To counteract such performative effect, it would be possible to rely on an interesting solution coming from the *Deckmyn* decision, where the reproduction and the transformation of a famous Belgium comic image by a far-right party to communicate a message judged discriminatory by the Court has been forbidden. The ECJ has considered (point 31) that "In those circumstances, holders of rights . . . have, in principle, a *legitimate interest* in ensuring that the work protected by copyright *is not associated* with such a message."[31] This intriguing solution could open interesting leeway for rights holders regarding deepfakes, even if it remains unclear after the *Deckmyn* case, on which grounds such actions could be brought. Their standing to sue could certainly be based on the harmful effect caused by the association of the protected content with a violent message that conflicts with essential democratic values, but would it also cover uses that are not criminal offenses but merely contrary to the intention of the artist as to the meaning of her creation?

Such a possibility would be comparable to the moral right to respect the "intellectual integrity" of the work or of the performance recognized, for example, in France. In an often-cited decision of the French Cour de cassation, the judges agreed with a famous singer Jean Ferrat who contested the possibility to use one of his songs in a compilation together with some pro-Nazi tunes.[32] According to them, "by ruling in this way, while an exploitation in the form of compilations with works by other performers being likely to alter their meaning, could not fall within the exclusive discretion of the assignee and required a special authorization from the artist, the Court of Appeal disregarded the above-mentioned text."[33] Under such an interpretation, if the deepfake has altered the meaning of a work or of a performance, the artist could oppose his moral right, even if the reproduction is materially correct and had been agreed upon by the holder of the economic right of reproduction.

2. The Need for Harmonization of Moral Right

There has been, so far, no harmonization of the moral right within the EU and no such project is pending. Historically, the moral right was considered outside the competence of the EU because of its non-economic and cultural dimensions. Member

30. *See* EU AI Act, *supra* note 2, art. 50.
31. *Deckmyn*, Case C-201/13 (2014).
32. Cour de Cassation [Cass.] [supreme court for judicial matters], oc., Feb. 8 2006, Bull. civ. V, No. 64 (Fr.).
33. *Id.*

States, due to their quite different approaches, were reluctant to any initiative in that respect. However, many reasons may now justify a move forward. First, there is already an indirect international harmonization of the moral right within the EU, through the adhesion of all Member States to the Berne convention (article 6 bis)[34] and more recently with the signature by the EU of the Beijing Treaty, albeit not yet ratified.[35] Second, the competence of the EU in the cultural field has been extended since the Treaty on the Functioning of the European Union (TFUE).[36] Third, the bizarre *Deckmyn* case shows that the ECJ is capable of a judicial harmonization even in the absence of any legal ground.

Adopting a moral right at the EU level would help the individual creators and performers to fight against undue intellectual appropriation of their works and performances, among other situations, in a deepfake. Such prerogatives already exist in various expressions within Member States, and it would not be difficult to demonstrate that misuse of the protected contents has a cross-border effect. Providing to the authors and performers the right to claim their quality when it is not traceable in AI output, or to oppose to the use of their name and quality; or offering them the possibility to oppose to the association of their creations or interpretations with offending content, as a form of violation of the right to respect them, would reduce the development of "mean" deepfakes. The French system—where moral rights are perpetual and not assignable—could be a valid source of inspiration in this perspective. Perpetuity would ensure that the prerogative would last as long as the risk of association subsists, even after economic rights have lapsed. As in France, a public entity could contest misuse and, by doing so, defend the honor and reputation of the dead authors or performers in the absence of heirs of the right holder, or whenever the heirs are acting contrary to the expressed will of the deceased person. The un-assignability of the moral right would limit the effects of the authorization given by the assignee of the economic right to generate deepfakes in contradiction with intention of the author or the performer.

34. *See* Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, as revised at Paris on July 24, 1971, and amended in 1979, 828 U.N.T.S. 221, at art. 6 bis.

35. Beijing Convention on Audiovisual Performances, June 24, 2012, 53 I.L.M. 1020, art. 5 (entered into force April 28, 2020) (EU is member):

> 1) Independently of a performer's economic rights, and even after the transfer of those rights, the performer shall, as regards his live performances or performances fixed in audiovisual fixations, have the right:
>
> (i) to claim to be identified as the performer of his performances, except where omission is dictated by the manner of the use of the performance; and
>
> (ii) to object to any distortion, mutilation or other modification of his performances that would be prejudicial to his reputation, taking due account of the nature of audiovisual fixations.

36. *See* Consolidated Version of the Treaty on the Functioning of the European Union (TFEU), art. 6, June 7, 2016, 2016 O.J. (C 202) 47 (recognizing that the EU's competences in the field of culture are to "carry out actions to support, coordinate or supplement the actions of the Member States.").

III. SHOULD WE EXTEND THE IP MODEL?

After weighing the pros and cons of an IP model for deepfakes (III.A), the Article concludes that the disadvantages of an extension of copyright or other exclusive IP right on one's likeness, voice or goods exceed the benefits of such a solution (III.B).

A. THE PROS AND CONS OF AN IP MODEL FOR DEEPFAKES

As demonstrated above, IP does not give comprehensive responses to three types of questions that should be taken into account when addressing deepfakes: the role of technology, the respect of the right holders' interests, and the guarantee of truth for the public.

1. Technological Challenge

IP law has a long history of adapting to technology. Even where its principles have remained stable, IP legislation has been constantly modified to integrate technological evolutions (radio, tv, satellite, digital, internet, etc.).Yet, so far, no new provisions have been adopted to deal with AI-generated content under EU law: No specific regime has been imagined to regulate the generation of outputs, including deepfakes. As current laws do not distinguish between different technological means, AI generation must respect IP rules whenever its process involves what may qualify as acts of exploitation under the law. The question is not so much theoretical but resides in the practical difficulty of proving the existence of such acts in a process where many operations remain unrevealed or are evanescent. Many initiatives are pending in Europe and in Member States to alleviate this burden by proposing presumptions of use. The starting point of this initiative is the observation that visible similarities may exist between AI-generated output and protected content, while acts of exploitation may be impossible to prove because of the complexity and the sheer number of operations required to generate the output.[37]

Assuming that this proof can be somehow demonstrated, there are still some "technological" exceptions that may hamper the possibility of exercising the exclusive right. Two sets of provisions are relevant: the exception for transient copy introduced in the Infosoc Directive in 2001, far before the AI-generated content explosion (article 5, paragraph 1), and the two exceptions dedicated to text and data mining (articles 3 and 4) in the DSM Directive, adopted in 2019.[38] These three provisions aim to facilitate a technical copy, being part of a wider process in which such copies have, in principle,

37. *See* FRENCH SENATE, CRÉATION ET IA: DE LA PRÉDATION AU PARTAGE DE LA VALEUR [CREATION AND AI: FROM PREDATION TO VALUE-SHARING] (July 9, 2025), Rapport d'information n° 842 (2024–2025) , https://www.senat.fr/rap/r24-842/r24-842.html [https://perma.cc/9Y22-YR6H]; *see also infra* notes 40 and 41.

38. Infosoc Directive, *supra* note 11, art. 5(1); Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC ["DSM Directive"], arts. 3–4. 2019 O.J. (L 130).

no specific economic value. Unlike the reproduction of a work doomed to be communicated, in whole or in part, to the public, these reproductions—due to the digital technology—are pre-conditions for the transmission or analysis of the data. These provisions, compulsory for Member States, have in common an aim to foster innovation by avoiding a too burdensome circuit of prior authorization. But while the transient copy exception is public order, the text and data mining exception for general uses (art. 4) can be reversed by an opt-out system, which forces the right holder to signal her opposition to the reproduction required to mine if she wants to retain her exclusive rights.[39] Consequently, the right holder can theoretically prevent reproduction made to "feed" an AI system by opting out preventively.

However, the concrete application of this provision demonstrates the many difficulties that right holders face when they exercise their opt-out option, especially when the content is displayed online. That is because only a mechanism readable by the machine is, then, opposable. In many cases, rights holders don't have the possibility to impose such mechanisms on online pages where their content is made available. Moreover, much of the scraping necessary to train AI models would have been done before the entry into force of this exception, which renders this late opt-out unnecessary. Finally, the opt-out is sometimes simply ignored, with reproduction made in spite of the right holder's opposition.

Right holders are currently looking for a way out of this dead end. Some settlements gave them hope that they could, at least, obtain money to compensate their loss. Some try to negotiate the past uses of their content by AI models or to sue them for infringement because they have been reproducing the content without their prior authorization before the entry into force of the DSM Directive, or in violation of the opt-out mechanism. International private law questions are evoked by tech companies to circumvent the territorial application of the directive, advocating that the training occurred in countries where the reproduction is *fair use*. The case law is still scarce and difficult to interpret at the EU level, in the absence of a decision of the ECJ.

Even if some voices have asked for a re-opening of the DSM Directive to better balance the interests of the right holders and the interests of tech companies in favor of the first ones, the Commission has so far shown no intention to do so. Nevertheless, some initiatives have been developed lately in certain Member States[40] and within the European Parliament[41] to encourage the adoption of presumptions which would reverse the burden of the proof, in the presence of certain elements, namely a resemblance between the output and pre-existing content. The mechanism is still unclear. Should it be a national competence based on procedural rules only? Should it

39. DSM Directive, *supra* note 38, art. 4.

40. CSPLA [SUPERIOR COUNCIL OF LITERARY & ARTISTIC PROPERTY], TASK FORCE REPORT ON RENUMERATION FOR CULTURAL CONTENT USED BY AI SYSTEMS—LEGAL COMPONENT (task force co-chair Alexandra Bensamoun & legal component author Joëlle Farachy) (June 25, 2025), https://www.culture.gouv.fr/Media/medias-creation-rapide/cspla-ai-cultural-content-remuneration-legal-component-english.pdf; FRENCH SENATE, *supra* note 37.

41. European Parliament Resolution of 10 March 2026 on Copyright and Generative Artificial Intelligence—Opportunities and Challenges, 2025/2058(INI), P10_TA(2026)0066.

be adopted at the EU level? Which type of event would trigger the presumption? As we have seen, the problem is even more difficult with regards to deepfakes, where the resemblances may not be linked to a specific work or performance but more globally to a part of the author's or performer's repertoire.

2. Interests of the Right Holder

The benefit of an intellectual property right is to provide to the right owner a legal exclusivity on exploitations of her protected content, so that she can shape the market ex-ante because her prior authorization is required to enter it. She has also the possibility to sue, ex-post, without having to demonstrate a harm different from the infringement to her monopoly.

Whether by refusing the authorization to reproduce and to display, or by suing the defendant for having exploited without permission, the right holder normally has the capacity to ban the publicity of deepfakes and therefore to protect herself against their potential deceptive effect and risk of substitution. The performer would for example oppose the reproduction of her performances to feed the AI system, so that no duplicate of her would be available to replace her. The trademark owner would also be capable of banning the use of the distinctive sign in commercials he does not want to be related to. The benefit of intellectual property rights also lies in the efficiency of their infringement procedures. The right holder does not need to demonstrate that he has suffered harm: Infringement is sufficient to trigger the action. However, she can also rely on robust procedural instruments harmonized by the 2004 Directive such as interim measures, information measures, or seizure.[42] Some procedures specific to the digital field, such as a notice and take down and stay down, have been developed. At the EU level, article 17 of the DSM Directive is forcing platforms to implement filtering measures to avoid the new presentation of a file, which has been signaled by the right holder as infringing his copyright or related right.[43] With such an instrument, the right holder can freeze the principle of non-liability of the hosting provider set in the E-commerce Directive since 2000 and reaffirmed in the Digital Services Act (DSA) and sue the platform for direct infringement if it does not implement correctly the appropriate measures of control.[44] He can force the platform to cooperate and to help him fighting against counterfeiting. This constitutes a procedural advantage as compared to the difficulties of private persons to convince platforms to withdraw other offensive contents.

42. Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights, 2004 O.J. (L 157).

43. DSM Directive, *supra* note 38, art. 17.

44. Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market ["E-commerce Directive"], 2000 O.J. (L 178); Regulation 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC ["DSA"], 2022 O.J. (L 277).

The exercise of the intellectual property right can also be positive: The right holder may decide to authorize deepfakes, potentially against remuneration. As the economic right is assignable or capable to be licensed, there is in principle no obstacle for such a market. As contractual law is not harmonized across the EU, only some provisions of the DSM Directive secure certain protections for authors and performers, with regards to remuneration, transparency, and revocation of the contract in case of inexecution (article 18 to 21 of the DSM Directive).[45] The performer may give her consent for digital duplicates that would appear in films instead of herself, the producer may authorized reproduction of AI-generated deepfakes where a song is interpreted with the voice of another performer: The possibilities are already numerous and will most likely continue to grow in the next few years. Nevertheless, the deceptive aspect of deepfakes reinforces the disturbing aspect of a market developing with the right holders' blessing so that they can earn royalties deriving from the exploitation thereof. This situation is made possible by the fact that, in most cases, IP rights don't protect the interest of the public in distinguishing truth from counterfeit, even if they usually leave space for freedom of expression.

3. IP Right and Truth: Two Non-Matching Concepts

The crux of intellectual property rights can be located somewhere between the private interest of the right holder and the public interest in the development of creation and innovation. The relevance of what truth is does not interfere with the exercise of the monopoly: Ultimately, truth is what the right holder decides it is. Even in the field of trademark where the guarantee of origin of the product is part of the essential function of the right, and where the assignment of the title shall be published accordingly, there is no legal obstacle to settle a trademark coexistence agreement with another trademark holder, even if such agreement is detrimental to the distinctiveness of the sign. The consumer may therefore be misled on the origin of the product, if the right holder decides so. Even if the absence of risk of confusion and prohibition of deceptive trademarks are key for the obtention of a trademark, this risk does not trigger any obligation to sue for the right holder, who can decide to leave counterfeiting goods on the market. For some right holders, the existence of such copies may be seen as an opportunity to gain some notoriety or to favor a secondary market for other categories of consumers. Such passive behavior, which may favor "fakes," is not considered abusive, although there is a possibility to lose the exclusive right by acquiescence and to legitimize the fake.

In the case of the creation of one performer's digital replica where the use has been agreed by the artist, there is no provision within the IP legislation which obliges the right holder to inform the public that the avatar displayed is not "real." Only the AI Act provides for such information, which must be disclosed, not by the artist, but by the deployer of the AI solution.[46] Other examples in the field of copyright demonstrate this

45. DSM Directive, *supra* note 38, arts. 18–21.
46. EU AI Act, *supra* note 2, art. 50 (transparency obligations for deployers of AI tools).

disregard for the truth: According to the moral right rules when they exist, the copyright holder can decide not to reveal the author's real identity and to publish his work anonymously or under a pseudonym; no mandatory duty to disclose the author's identity exists. It's even the opposite as, according to some legislation, such as the French one, the unconsented revelation of the authorship may constitute a violation of the moral right to attribution and consequently, a copyright infringement.[47] The right to integrity also is exercised by the author or her heirs as regards what *they* believe is betraying the work and not as a duty to inform the public of the fake.[48]

Finally, the concept of truth is a rather complicated one in the field of artistic creation, where most works are fictional, except for press and documentary which report information. Would a deepfake of a video game be less "true" than the model from which it has been generated? If an actress performs a scene of a movie in a deepfake, isn't that as fictional as the same actress performing a scene in a "real" movie? If a performer has accepted the AI processing of her image and voice, and eventually supervises the result for a dangerous scene of a movie, is it really different from the same scene being shot with a stuntman? The AI Act itself, in the final sentence of article 50 paragraph 4, foresees the risk of a systematic obligation of disclosure of the existence of generated or manipulated content in such fictional creations, assuming that the information upon the algorithmic origin of the element shall be revealed in an appropriate manner that does not hamper the display or enjoyment of the work.[49] It appears that IP is not an appropriate instrument to secure the liability of the message conveyed by the protected subject matter.[50] The economic dimension of these assets supersedes the public interest to value the truthfulness of this message. This is even more true concerning the transformation by deepfakes of works that are already fictional and from which the expectancies of the public are not related to the reality.

Thus, the EU IP system seems adapted to tackle the problems raised by deepfakes, but only partially. It may, to a certain extent, provide opportunities for the right holder to control the use of the protected content whenever it is still identifiable: by exercising the monopoly in a positive manner or by bringing defensive action against their misappropriation in deepfakes and enforcing his rights against platforms through notice and take down and stay down or filtering mechanisms. From a perspective of freedom of expression, the exception of parody can give to the deployers a certain margin of maneuver to produce AI-generated funny or critical deepfakes as far as they don't associate the content with offending messages. Yet, in the absence of harmonization of the adaptation right and of the moral right at the EU level there is no

47. CODE DE LA PROPRIÉTÉ INTELLECTUELLE [Intellectual Property Code] art. L. 121-1 (Fr.).

48. Victor Hugo's heirs have had, for example, different views on what should be the respect for the writer's monumental work: They sued an author who had created sequels to *Les Misérables* starting from key characters in the novel, namely because Javert, who had died in Hugo's version, was living again, whereas they did not object to the Disney cartoon *The Hunchback of Notre Dame*, based on Hugo's 1831 novel *Notre-Dame de Paris*, despite the fact that all the main characters are dead at the end of the novel.

49. EU AI Act, *supra* note 2, art. 50(4).

50. Even with respect to the press publishers' right, whose adoption was namely justified by the need to combat fake news, no provision effectively imposes such an obligation.

EU common ground on which the author or the performer can rely to contest the alteration of the content in the deepfake. Furthermore, when the deepfake bears no resemblance to the sources it was produced from, the mechanism of IP proves difficult to apply. The current discussion on the necessity to introduce presumptions to reverse the difficult demonstration of the existence of an act of exploitation may render the system more efficient. The extent to which exceptions may apply and the practicability to opt out are heavily disputed. Finally, with a variation depending on the IP right concerned, the system is offering a poor guarantee to the public on the faithful character of the AI-generated content. No transparency obligation is imposed to the right holder to reveal the fakeness of an output: On the contrary, IP's monopolistic structure would help "launder" the deepfake by the mere consent of the right holder. This overall limited enthusiastic assessment prevents the promotion of the extension of IP mechanisms to the protection of likeness, voice, or goods on behalf of everybody.

B. AN INAPPROPRIATE EXTENSION OF INTELLECTUAL PROPERTY TO ONE'S LIKENESS, VOICE OR GOODS

The current debates around the No Fakes Act in the United States and of the revision of the copyright law in Denmark raise the question of creating new IP rights to one's likeness or voice at the EU level.[51] The numerous scams, sextortions, and revenge porn operated with the use of deepfakes have raised a growing concern within the public, so that people are willing to be "protected" against such damages. But as we have seen, most of these questions have already been addressed by the criminal rules at the national level, such as the recent Italian law.[52] Therefore, the rationale for creating such monopolies should be different from these public order rules and would consist either of the individual bringing a civil action against the exploitation of the deepfake or of steering revenues from this exploitation by assigning the right against economic counterparts. Namely, Denmark is pleading for such an extension at the EU level, arguing that the enforcement of IP rights by platforms is more efficient than the responses given in case of the violation of personality right or unfair competition.

This argument seems fallacious, in the first place because the enforcement efficiency justification can be easily rebutted with the many difficulties encountered by copyright holders depending on the very unstable strategy of the platforms. The second reason why this extension of intellectual property to one's likeness or voice seems irrelevant resides in the danger of creating such legal monopolies without any social counterpart. IP rights have been implemented because they enhance creativity, encourage innovation for the benefit of society as a whole, and limit the reproduction of immaterial goods. Such expectations justify that some restrictions are opposed to other

51. *See* NO FAKES Act of 2025, S. 1367, 119th Cong. (2025); *supra* note 9 and accompanying text.
52. *See* Art. 612-quarter (It.), *supra* note 8.

fundamental rights. In the case of one's likeness or voice, there is no substantial effort or investment made by the person to create them and no reward shall be expected. They just are. The professor is not playing a role in a creation while teaching his class to the pupils while being recorded for a documentary film, as it has been decided by the Cour de cassation in a famous case about the documentary *Etre et avoir* (*To Be and To Have*), nor can the pupils be considered as performers when they are just being filmed in their day-to-day life.[53]

Such an effort may exist for certain categories of persons—such as models, singers, athletes, influencers, reality TV stars—for whom one's image has an economic value, because of their notoriety. While the introduction of a right of publicity under the form of an exclusive right at the EU level may be worth discussing, for the average person, being vested with an intellectual property right will be unnecessary and potentially dangerous. The multiplication of these monopolies may threaten fundamental freedoms to a far-reaching extent. Who would take a photo of an event of ten thousand attendees if each of them can claim an intellectual property right on his image? Who can afford to pay royalties to register the voices of the children singing in a kindergarten? Monopolizing these elements will just lead to a further commodification of the human beings and excite greed: One can imagine a father "selling" the likeness of his kids, or people trading the image of a deceased heir. Rivalry on the ownership may be dramatic: Who should have the exclusive right on the image of his heir? Finally, many contracts signed to use social networks already encompasses clauses where the user assigns a non-exclusive unlimited license to exploit all the elements posted, even after the user resigns. Consequently, even with a new IP right, the social network would be entitled to develop avatars of their members without any further authorization and against no remuneration.

The propensity to analyze any relationship in terms of property may finally lead to weird solutions. Shall I be considered as owning my voice or features, when the images and capture of my voice are being recorded by someone else? So far, IP rights lie with the photographer, not with the subject of the photograph. How will those two pretentions be combined? A similar question was raised about the ownership of the image of a tangible good. Shall the owner of the good be recognized as the owner of all potential images of this good? The French Cour de cassation, after several hesitations, wisely decided not to provide exclusive rights to the owner of a tangible good (here a mansion in Rouen) on the photograph thereof and to submit the action of the latter to the demonstration of the harm he suffered from the use of this image.[54]

Additionally, as demonstrated above, if consent launders the fakeness of the output, granting IP rights to persons on their likeness or vocal expression would not further help the public identify whether the AI-generated output is a fake or not. The proposed extension would be detrimental to the society, by multiplying the risk of conflicts between these millions of right holders claiming exclusivity on their likeness, without

53. Cour de cassation [Cass.] [supreme court for judicial matters] 1e civ., Nov. 13, 2008, Bull. civ. I, No. 259 (Fr.).

54. Cass. ass. plén., May 7, 2004, Hôtel de Girancourt, Bull. civ. n° 516 (Fr.).

providing any benefit in terms of innovation or creation and very tentative chances of incomes for the right holders. Besides, it is uncertain whether creating IP rights is necessary to combat unwelcome uses of deepfakes, as other means of redress exist without the need to implement such monopolies.

IV. ALTERNATIVES FOR "INDIVIDUAL" DEEPFAKES

Other solutions already exist at the national level or across the EU. There is a wide spectrum of possibilities to regain control over deepfakes. They range from the most comprehensive, the ex-ante exercise of personality right, to a mere ex-post action, based on tort law and demonstration of the harm. This section will only develop the first one under two types of legal frameworks: the still unharmonized personality rights (IV.A) and the EU-wide protection of personal data (IV.B).

A. PROTECTING INDIVIDUALS THROUGH PERSONALITY RIGHTS

Natural persons benefit in most EU countries from a protection against the violation of their privacy, and Courts or the EU legal framework have in practice sometimes recognized an embryo of right of publicity.[55] The extreme variety of the situations and legal regimes within the different Member States cannot be exposed in this paper but demonstrates the lack of uniformity of the solution at the EU level.[56] The French system mostly relies on case law based on the article 9 of the Civil Code, which protects the right to privacy.[57] The courts have progressively built a comprehensive protection of the right to one's image, including in some situations the protection of its commercial value for notorious persons, but there is no such thing as an autonomous right like in Spain, where the right to one's image is seen as an autonomous personal right, independent from the right of honor and the right to privacy and is laid out by the Constitution and the Organic Law of May 5, 1982.[58] The Italian system distinguishes between typical rights in the legal statute and un-numerated deriving from the case law.[59] Therefore, in Italy, the defense of the image, likeness, or reputation derives from

55. *See* Jan Klink, *50 Years of Publicity Rights in the United States and the Never Ending Hassle with Intellectual Property and Personality Rights in Europe*, 4 INTELL. PROP. Q. 363, 364 (2003); Daniel Gervais & Martin L. Holmes, *Fame, Property, and Identity: the Scope and Purpose of the Right of Publicity*, 25 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 181, 186 (2014).

56. Tatiana Synodinou, *Image Right and Copyright Law in Europe: Divergences and Convergences*, 3 LAWS 181, 183–89 (2014); Kateryna. Moskalenko, T*he Right of Publicity in the USA, the EU, and Ukraine*, 1 INT'L COMPAR. JURIS.113, 115–16 (2015).

57. *See* CODE CIVIL [CIVIL CODE] art. 9 (Fr.).

58. *See* CONSTITUCIÓN ESPAÑOLA art. 18 (Spain) (constitutional provision); Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen [Organic Law of May 5, 1982], B.O.E. 1982, 11196 (Spain) (statute implementing the personality rights protected under the constitution).

59. *See* Giorgio Pino, *The Right to Personal Identity in Italian Private Law*, *in* THE HARMONIZATION OF PRIVATE LAW IN EUROPE 225, 227 (M. Van Hoecke & F. Ost, eds., 2000) ("It is easy to conclude, then, that according to this traditional position, the Civil Code regulates only patrimonial-economic relations, while

a combination of privacy right, extension of copyright and tort law. In Germany, the federal Supreme Court has developed a general, non-transferable, right of personality, based on articles 1 and 2 of the German Constitution and section 823 of the German Civil Code.[60]

There is currently no real consensus among EU Member States on the harmonization of such personal rights or on the precise scope of protection that individuals should enjoy. Do these rights protect only an individual's image, or do they also cover her voice, general likeness, and reputation? Can such rights be licensed or assigned? Do they survive the death of the person and if so, who would be entitled to exercise them post-mortem? As far as personality rights are concerned, deepfakes are therefore addressed primarily at the national level. This may generate significant difficulties when action must be taken against a platform. For instance, would it be possible to ask for the removal of a deepfake with the likeness of a dead person in France, where personal rights normally expire upon death, on the ground that it violates an exclusive license to use this likeness under Italian law? Greater uniformity would certainly be welcome, even if EU law already offers substantial protection to personal likeness and voice when they are subject to data processing, as is generally the case for AI-generated deepfakes.

B. PROTECTING PERSONAL DATA

European law has a vibrant set of rules, namely the famous GDPR, which protect the personal data of its citizens.[61] This paper is not the place to analyze the details of this regulation, but it is sufficient to recall that image, voice, and name are considered as protected personal data, whenever they can serve to identify an individual. This vests in the person a bundle of rights regarding the data processing that can be applied uniformly throughout the EU and can be used against platforms when they operate on this territory, notwithstanding the location of their headquarters. The ECJ, considering that personal data is an autonomous notion of EU law, has rendered many decisions providing for a common interpretation of this notion, and of its regime.[62] If the consent is not always required for data processing because other grounds can legitimate such processing, the GDPR may nevertheless offers a standing for action against the entity which would have used personal likeness of one's person in many cases. For example, a deepfake can be forbidden if it has been processed in an algorithm that performs facial

the moral development of the human being is committed to constitutional law, and protected only by penal law.").

60. Grundgesetz [Basic Law], arts. 1–2; Bürgerliches Gesetzbuch [Civil Code] § 823 (Ger.).

61. Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) [hereinafter "GDPR"], 2016 O.J. (L 119) 1.

62. *See* Eur. Data Protection Supervisor (EDPS) v. Single Resolution Bd. (SRB), Case C-413/23 P, ECLI:EU:C:2025:645 (First Chamber Sep. 4, 2025).

recognition. It also prohibits the manipulation of the content and the processing, without the consent of the person, of data regarding race, health, sexual preference, or political beliefs. In any case, the entity which processes the personal data must comply with many requirements of the GDPR such as transparency, data minimization and security, and respect of the purpose of the processing. Whether these conditions can be met at all when creating a deepfake is not evident, as the GDPR doesn't have any specific exception for parody.

The GDPR does not confer a systematic right to authorize data processing but establishes other rights that may be very useful to oppose to the public display of a deepfake, such as the right to information and the right to be forgotten.[63] With respect of the right to information, the creator of the deepfake must inform the individual depicted that she is creating a deepfake of her person. In regard to the right to be forgotten, the controller, who can be either the maker of the deepfake or the platform publicly diffusing it, must delete personal data if the subject requests this. While the GDPR was not conceived with the issue of deepfakes in mind, its broad scope of application is capable of covering some of the issues pertaining to false representation of persons. However, and similarly to some countries with regards to personality rights, it remains true that the GDPR does not cover the scenario of a death, so other solutions may be necessary to adopt.

63. GDPR, *supra* note 61, arts. 13–14 (right to be informed), art. 17 (right to be forgotten).

V. CONCLUSION

This Article has sought to demonstrate that there is currently no lack of legal protection against the harmful consequences of deepfakes in Europe—quite the contrary. The existing rules provide sufficient grounds for persons whose likeness or creations have been commercially exploited to share the benefits of that exploitation, provided they are willing to consent to it, either on the basis of European IP rules whenever these elements are subject to protection, or according to national provisions regarding personality rights. Consequently, the path forward currently promoted by the Danish presidency and inspired by the debate around the No Fakes Act to create an extension of copyright to everyone's likeness, appears not only unnecessary but also hazardous. Such a development would be dangerous for the rationales of the intellectual property system as it would grant monopolies without social counterparts. It would be detrimental to the protection of individuals by multiplying opportunities for commodification that would ultimately benefit powerful economic actors, who already impose on individuals an unremunerated licensing of these immaterial assets through unbalanced contracts. Nor would it increase the public's ability to recognize and identify the deceptive character of the output.

Should reforms be initiated at the EU level, three different solutions could be explored:

(1) strengthening the obligations of the platform to monitor the content they contribute to diffuse and to quickly and efficiently withdraw litigious content when deepfakes are involved, without hindering the freedom of creation and political critique; the DSA and other specific texts are already offering such solutions, the application of which must be rendered more efficient;
(2) alleviating the burden of the holders of intellectual property right to prove that AI-generated content, which somehow evokes protected content, like deepfakes generally do, implies an act of exploitation covered by the monopoly, particularly when there is a risk for the public to associate the deepfake with the right holder; this could be achieved by procedural presumption but also by the harmonization of the adaptation right and of the moral right;
(3) creating a specific regime for the deceased person, which would be more grounded on public interest to the respect of the memory of the deaths than on the private interests of some heirs, not as concerned by this need for respect as they may be by their own assets.

Tort Law Protections for Individuals' Images in Commonwealth Jurisdictions

Graeme W. Austin*

* Chair in Private Law, Victoria University of Wellington; Professor of Law, University of Melbourne. Thanks to Qadir Qeidary and Lulu Shepherd for excellent research assistance, and to the staff of the Columbia Journal of Law & Arts for an expert and careful edit.

Introduction

> "[U]nder English law it is not possible for a celebrity to claim a monopoly in his or her image, as if it were a trademark or brand." Lord Walker of Gestingthorpe in *OBG Ltd. v. Allan* [2008] 1 AC 1 [293].

The rapid development of AI-facilitated deepfake technology has provoked new interest in the availability and efficacy of legal protections for individuals' images and other aspects of their personae.[1] The U.S. Copyright Office has concluded that existing U.S. laws "fail to provide fully adequate protection."[2] The "NO FAKES" bill, introduced to the U.S. Congress in 2025, aims to address at least some of these gaps.[3] This Article adds to the discussion that has been provoked by deepfake technology by surveying common law tort law protections for image rights in Commonwealth jurisdictions.[4]

Two themes emerge from this survey. First, notwithstanding the absence of a distinct tort of misappropriation of personal identity[5] in some Commonwealth

1. *See* Jane C. Ginsburg & Graeme W. Austin, *Deepfakes in Domestic and International Perspective*, 48 COLUM. J.L. & ARTS 297, 300–10 (2025).

2. U.S. COPYRIGHT OFF., COPYRIGHT AND ARTIFICIAL INTELLIGENCE PART 1: DIGITAL REPLICAS 22–23 (July 2024), https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf [https://web.archive.org/web/20260226153447/https://copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf].

3. NO FAKES Act of 2025, H.R. 2794, 119th Cong. (2025) (as introduced April 9, 2025).

4. The following discussion is limited to private law protections. It does not consider the increasing array of criminal provisions directed at deepfakes. In some jurisdictions, consumer protection laws prohibit unlicensed commercial uses of an individual's image. For an example from Australia, see Talmax Pty Ltd. v. Telstra Corp Ltd. [1997] 2 Qd R 444 (Supreme Court of Appeal), applying s 52 of the Trade Practices Act 1974 (Cth). *See also* Competition and Consumer Act 2010 (Cth) sch 2 (Austl.). Elsewhere, advertising codes prohibit conduct of this kind. Singapore's Code of Advertising Practice (2008), for example, includes prohibitions against portraying or referring to any person without permission. And, relevant to the question of deepfakes, the Code provides that advertisements and sales promotions "should not manipulate (such as through electronic morphing) any person . . . to create a misleading or untruthful presentation." SINGAPORE CODE OF ADVERTISING PRACTICE, c. 3, §§ 13.1, 13.3 (3d ed. Feb. 2008).

5. In this paper the terms "misappropriation of personal identity" and "right of publicity" are used interchangeably. For present purposes, they refer to legal claims in respect of unlicenced use of aspects of another's identity, such as physical appearance or voice. The approach is consistent with that in some of the case law discussed below. For example, in Anil Kapoor v. Simply Life India, 2023 SCC OnLine Del 6914 [25] (High Court of Delhi), the plaintiff was seeking protection for "his personality rights, publicity rights and elements associated with his persona, such as: his name; his voice; his photograph/image/likeness; his manner of speaking and dialogue delivery; his gestures; his signatures, etc." The task of delineating the conceptual justifications for protections of image rights, and the intertwining of privacy and commercial

jurisdictions—a position reflected in Lord Walker's observations about the law of England and Wales—the tendency overall has been toward enhancing legal rights for individuals to exploit and control uses of their images.[6] The evolution of other private law causes of action, including defamation, breach of confidence, and passing off, has gone some way toward filling this gap. Even so, as the discussion below outlines, the protections are piecemeal and incomplete.

Second, in Commonwealth jurisdictions in which a distinct tort of misappropriation of personality *is* recognized—this Article considers case law from Canada, South Africa, and India—the private law analysis is coming to be infused with constitutionally-mandated commitments to protections for individual dignity.[7] We see in these jurisdictions an alignment between the decisional law and academic commentary characterizing misappropriation of an individual's image as dignitary affront.[8] The strands of authority are thin, but these developments appear to align with the international movement toward a greater focus on human rights in the development of domestic private law jurisprudence.[9] A full exposition of this development is beyond the scope of this Article: For present purposes, it suffices to suggest that the protection of image rights is another frontier in which rights discourse is infusing private law.[10] As is discussed below, courts in India have already deployed these ideas in cases involving deepfakes of Bollywood movie stars.[11] Going forward, recognition of human dignity as an organizing principle in the context of misappropriation of individuals' images could bolster the view—one reflected in the NO FAKES bill—that protections against deepfakes should not be confined to celebrities.[12] The new intellectual property right to be created by this bill would vest not only in celebrities seeking to monetize (or preserve the ability to monetize) attributes of their personalities, but also, as one of the bill's sponsors noted, in "everyday" individuals.[13]

---

concerns is insightfully discussed in JENNIFER E. ROTHMAN, THE RIGHT OF PUBLICITY: PRIVACY REIMAGINED FOR A PUBLIC WORLD (2018).

6. Adam Speker & Lily Walker-Parr, *Copyright, Moral Rights, and the Right to One's Image*, *in* THE LAW OF PRIVACY AND THE MEDIA 401, 430 (Nicole A. Moreham & Adam Speker eds., 4th ed. 2024).

7. For an influential analysis of the different senses in which the law protects human dignity, see Jeremy Waldron, *How the Law Protects Dignity*, 71 CAMBRIDGE L.J. 200 (2012).

8. *See, e.g.*, Jeffrey Rosen, THE UNWANTED GAZE: THE DESTRUCTION OF PRIVACY IN AMERICA 19 (2000).

9. *See, e.g.*, François du Bois, *Private Law in the Age of Rights, in* PRIVATE LAW AND HUMAN RIGHTS: BRINGING RIGHTS HOME IN SCOTLAND AND SOUTH AFRICA 12, 13 (Elspeth Reid & Daniel Visser eds., 2013).

10. The development is unsurprising, given the close jurisprudential links between privacy and image rights. *See, e.g.*, Jonathan Morgan, *Privacy, Confidence and Horizontal Effect,* 62 CAMBRIDGE L.J. 444 , 445–50 (2003).

11. Anil Kapoor v. Simply Life India, 2023 SCC OnLine Del 6914 (High Court of Delhi); Aishwarya Rai Bachchan v. Aishwaryaworld, 2025 SCC OnLine Del 5943 (High Court of Delhi).

12. NO FAKES Act, *supra* note 3.

13. *See* Press Release, Rep. Maria Elvira Salazar, Salazar Introduces the NO FAKES Act (Sept. 12, 2024), https://salazar.house.gov/media/press-releases/salazar-introduces-no-fakes-act [https://web.archive.org/web/20260201214853/https://salazar.house.gov/media/press-releases/salazar-introduces-no-fakes-act]. The relevant part of the statement began: "From the biggest entertainers to everyday Americans, non-consensual voice and image clones can ruin careers, deceive families and friends,

Celebrities have no monopoly on the dignitary harms deepfakes can inflict (even if the commercial implications might differ), perhaps especially so in the sexualized contexts in which deepfakes are prevailingly created and disseminated.[14] The invocation in Commonwealth jurisdictions of dignitary principles in right of publicity cases might advance our understanding of relevant interests and harms of deepfakes, and appropriate legal responses to them.[15]

I. IDENTITY PROTECTIONS WITHOUT PUBLICITY RIGHTS

A range of other common law torts protect at least some attributes of personality. Surveyed here are defamation, breach of confidence, and passing off. These liability theories are principally concerned with interests in reputation, the ability to control dissemination of confidential information, and protecting commercial goodwill. Even so, they have served to provide some claimants with remedies against misappropriation and injury to their image.

A. DEFAMATION

The famous 1931 English case of *Tolley v. J. S. Fry and Sons Ltd.* offers us an insight on how far the commercialization of celebrity has developed during the past decades.[16] Tolley was one of England's most famous golfers, the defendant a prominent chocolate manufacturer. Without authorization, the defendant used a caricature of the plaintiff in its advertising, depicting him with one of the defendant's chocolate bars protruding from his pocket. In the illustration, the plaintiff was accompanied by a golf caddy; in the text accompanying the illustration, the caddy likened the excellence of the defendant's chocolate to that of the plaintiff's drive. An amateur sportsman, Tolley based his defamation claim on the allegation that the defendant's depiction of him "meant . . . that the plaintiff had agreed or permitted his portrait to be exhibited for the purpose of the advertisement of the defendants' chocolate" and that he had "done so for gain and reward."[17] The implication was that "he had prostituted his reputation as an

and traumatize victims." As currently drafted, there are no geographical limitations on the individuals in whom the new intellectual property rights are vested. Both Americans and foreigners would benefit from the rights.

14. *See* Press Release, David Chiu, City Att'y, San Francisco, City Attorney Sues Most-Visited Websites that Create Nonconsensual Deepfake Pornography (Aug. 15, 2024), https://sfcityattorney.org/city-attorney-sues-most-visited-websites-that-create-nonconsensual-deepfake-pornography/ [https://web.archive.org/web/20260222142633/https://sfcityattorney.org/city-attorney-sues-most-visited-websites-that-create-nonconsensual-deepfake-pornography/].

15. The conceptual foundations for the right of publicity have long been debated. For a comprehensive survey of different perspectives in common and civil law jurisdiction, see HUW BEVERLEY-SMITH, ANSGAR OHLY & AGNÈS LUCAS-SCHLOETTER, PRIVACY, PROPERTY AND PERSONALITY: CIVIL LAW PERSPECTIVES ON COMMERCIAL APPROPRIATION (2005). For an insightful analysis in the deepfake context of the dignitary interests at stake in right of publicity claims, see Michael P. Goodyear, *Dignity and Deepfakes*, 57 ARIZ. ST. L.J. 931, 939–57 (2025).

16. [1931] AC 333 (HL).

17. *Id.* at 337.

amateur golf player for advertising purposes," and that he had "been guilty of conduct unworthy of his status as an amateur golfer."[18] Tolley led evidence that the belief that he had exploited his image commercially might cause him to be called upon to resign his membership of any reputable club. Overturning a divided Court of Appeal,[19] the House of Lords held the advertisement to be defamatory.

It is not that celebrity endorsements were unknown at this time. Celebrities have featured in English advertising since at least the 1880s, when Lillie Langtry became the face of the Pears soap brand.[20] Given the current prevalence, even ubiquity, of celebrity endorsements, including by prominent sportspeople, we might think that *Tolley v. Fry* is an antiquarian artifact of a simpler, less commercialized past—even if, at the time, the potential for sullying the plaintiff's image by associating him with a commercial product provided the basis for establishing the defamatory meaning. But the continued relevance of the decision, at least in some contexts, is suggested by a 1995 decision of the High Court of Singapore involving a prominent Member of the Singaporean Parliament, whose photograph was used without authorization in advertising for a restaurant.[21] The plaintiff was also an advocate and solicitor. Here, the defamatory meaning conveyed by the photograph was that the plaintiff had consented to the use of his image for personal gain or to sponsor a private restaurant "and that he had done so by taking advantage of his position as a Member of Parliament and also for the benefit of promoting himself as an advocate and solicitor."[22] In the court's view, the defamatory sting was worse than in *Tolley v. Fry*, "as the tone and standards of any holder of high public office and those of an advocate and solicitor required a careful, proportionate and purposeful expression in the law of defamation."[23] Reflecting expectations as to fastidiousness required of Singaporean politicians, the judge observed that the imperative of protecting an individual's good name applied "a fortiori" in the case of a holder of public office.[24]

The defamation tort has also been successfully applied in cases involving the unauthorized use of an individual's image in sexualized contexts. For example, a successful Singaporean fashion model was found to have been defamed when an escort agency used her photograph in Yellow Pages advertising.[25] And a prominent Australian rugby league player succeeded in defamation proceedings against an Australian

18. *Id.* at 133.

19. Tolley v. J S Fry and Sons Ltd. [1930] 1 KB 467 CA.

20. Joey Sanders, *Pens and Soap: Comparing British Advertising from the Victorian Era Through Historiographic and Female Lenses*, 9 GEN.: BROCK UNIV. UNDERGRADUATE J. HIST. 80, 85–6 (2024).

21. Chiam See Tong v. Xin Zhang Jiang Rest. Pte Ltd. [1995] 3 SLR 196 (Sing.). The advertisement appeared in the *New Paper*, at the time Singapore's second-highest circulating newspaper. The Chinese and English language versions of the handbills differed. The former explicitly referenced the charitable event, but the latter did not. The case focused on the meaning conveyed by the English-language version.

22. *Id.* at 199.

23. *Id.*

24. *Id.*

25. Hanis Saini Hussey v. Integrated Info. Pte Ltd. [1998] SGHC 219 (Sing.). The plaintiff's original claim was against the publisher of the advertisement. It joined the operator of the escort agency to the proceedings. After the publisher settled, it sought indemnification from the agency.

magazine that had published a photograph of him in a dressing room shower in which his genitals could be seen. Accompanying the photograph was text "of a suggestive nature, emphasizing a display of masculine nudity."[26] Here, the defamatory sting was that the plaintiff had deliberately permitted such a photograph to be published. In addition, the publication would cause members of the public to think that he was unfit to continue with the promotional work with young people with which he had been engaged.[27]

Protections afforded by defamation can sometimes go further than torts protections for commercial interests in celebrity, as a recent South African case illustrates.[28] A twelve-year old girl was photographed on a surf beach by a freelance photographer who then sold the photograph to the defendant publisher. The publisher featured the photograph on the cover of a surfing magazine, accompanied by the words "filth," "all-natural Eastern Cape honey," and "100% pure filth photos inside."[29] As glossaries of surfing slang explain,[30] the word "filth" can bear positive connotations in "surf lingo."[31] The South African court was not, however, convinced as to the defendant's evidence on the point, and upheld the defamation claim.[32] As is discussed below, the court indicated it might also have found for the plaintiff based on the misappropriation of the plaintiff's image.[33]

At the same time, there may be doctrinal limitations to the capacity of the defamation tort to vindicate image rights, even in sexualized depictions. A well-known decision of the House of Lords, *Charleston v. News Group Newspapers Ltd.*[34] was a

26. Australian Consol Press Ltd. v. Ettingshausen [1993] NSWCA 10, 2 (Austl.). The case went to the Supreme Court of New South Wales Court of Appeal on the question of whether damages were excessive.

27. *Id.* Referring to Haelan Labs., Inc. v. Topps Chewing Gum, Inc., 202 F.2d 866, 868 (2d Cir. 1953), Kirby P. (the president of the appellate bench) observed that the defamation tort here was providing relief for harms that would be compensable under U.S. state law protections for misappropriation of personality. But, in his view, the absence in New South Wales common law of a tort directed at misappropriation of personality should mean that damages in successful defamation claims should be kept in check, and limited to compensating for the harm to the plaintiff's reputation. He observed: "Other notions, which arguably should or might in other places, or from a general sense of fairness ought to give rise to an entitlement to damages, must be ignored. They are irrelevant to our law." *Id.* at 10.

28. Wells v. Atoll Media (Pty) Ltd. 2010 (4) All SA 548 (WCC) at ¶ 51 (S. Afr.).

29. There was evidence that the photograph had been used as a pin-up photograph at an all-boys High School, and the girl herself had received social media postings describing her as a "slut" and a "little porno star." *Id.* at para. 5, 9.

30. *How to Speak the Surf Lingo?*, SURF & SUN (Aug. 16, 2022), https://www.surfandsun.com.au/blog/2022/how-to-speak-the-surf-lingo.php [https://web.archive.org/web/20260213194641/https://www.surfandsun.com.au/blog/2022/how-to-speak-the-surf-lingo.php].

31. *Wells*, (4) All SA 548 (WCC) at para. 10.

32. It should not be thought, however, that defamation can always achieve protections for the image of non-celebrities—the "everyday" individuals whose interests the sponsors of the NO FAKES Bill purport to be solicitous. The court's analysis suggests that, with a little more care by the defendant —either in its choice of terms that accompanied the photograph, or in its preparation of evidence about the likely understanding of "filth" to readers of surfing magazine—it is possible that the defamatory claim would not be made out.

33. See *infra* Part II.

34. Charleston v. News Grp. Newspapers Ltd. [1995] 2 AC 65 (appeal taken from Eng.).

precursor to current controversies involving sexualized deepfakes of famous individuals. The two plaintiffs were leading actors in an astonishingly successful television soap opera, *Neighbours*. An English tabloid published an article featuring the actors' faces superimposed onto near-naked bodies, apparently engaging in, as Lord Bridge of Harwich described it, "an act of intercourse or sodomy." One headline read: "Porn Shocker for Neighbours Stars."[35] None of this was done with the consent of the plaintiffs. Even so, the defamation claim failed. An analogue version of clickbait, the text of the newspaper article clarified that plaintiffs' faces had been superimposed on the bodies of "real porn models" by the makers of a pornographic computer game.[36] As described by Lord Bridge, the article combined eye-catching prurience with "self-righteous indignation" directed at the makers of the game,[37] which contrasted "oddly with the prominence given to the main photograph."[38] Nonetheless, the House of Lords held that a reasonable reader must be assumed to have read the whole article, including the less prominent section that explained the provenance of the images, thereby removing the defamatory sting.[39]

The orthodox, and largely uncontroversial, position that defamation claims do not survive the death of the plaintiff is a further constraint on the defamation tort as a vehicle for protecting image rights.[40] The point arose in the image rights context in a 2023 decision of the High Court of Delhi concerning a famous Bollywood actor who had died under circumstances then regarded as "suspicious."[41] His father sought to enjoin the use of his son's name, caricature, lifestyle or likeness in forthcoming films and other ventures, alleging that "gross . . . damage" would be done to the actor's "vast reputation/goodwill amongst the public across the world and the Hindi film industry" by associating with him with an "immoral" and "promiscuous" feature film.[42] The claim was summarily dismissed as misconceived, partly on the basis that the defamation tort does not survive the death of the plaintiff.[43] In contrast, and in line with U.S.

35. *Id.* at 69.

36. *Id.*

37. *Id.*

38. *Id.* The role of the bane and antidote principle in Anglo defamation law is broadly analogous to the effect of disclaimers where liability is based on trademark or unfair competition principles. *See id.* at 313, 316–17; HOWARD S. HOGAN & STEPHEN W. FEINGOLD, INTELLECTUAL PROPERTY LAW IN CYBERSPACE 425 (G. Peter Albert, Jr. & Intell. Prop. L. Assoc. 2d ed. 2011).

39. *Charleston*, 2 AC (HL) at 72.

40. In New Zealand, this question has been raised in the context of a decision of the New Zealand Supreme Court holding that a criminal appeal may be pursued after the death of the convicted person. Ellis v. The King [2022] NZSC 114. In its analysis, the apex New Zealand court referred to the law of the indigenous Māori (tikanga), according to which a person's reputation endures after death. While the current New Zealand law does not allow a civil claim for defamation to be brought after the death of the plaintiff, scholars have questioned whether this position aligns with tikanga principles. *See, e.g.,* Pete McKenzie, *The Afterlife of Peter Ellis*, NORTH & SOUTH (Mar. 16, 2021), https://northandsouth.co.nz/2021/07/27/peter-ellis-tikanga/ [https://web.archive.org/web/20260213194147/https://northandsouth.co.nz/2021/07/27/peter-ellis-tikanga/].

41. Krishna Kishore Singh v. Sarla A. Saraogi, 2023 SCC OnLine Del 3997 (High Court of Delhi).

42. *Id.* at 18.

43. *Id.* at 35. A further ground for dismissing the case was that it was speculative. No clear information was before the court about the content of any proposed depictions of the actor's death.

protections for post-mortem image rights under some states' laws,[44] the NO FAKES bill would protect rights in digital replicas after death.[45]

B. BREACH OF CONFIDENCE AND TORTS PROTECTIONS FOR PRIVACY

Lord Walker's observation that English law does not protect rights of publicity was made in the context of a claim for breach of confidence involving one of the most storied celebrity weddings in recent decades: between actors Michael Douglas and Catherine Zeta-Jones.[46] The couple had sold the exclusive right to publish photographs of their wedding to *OK!* magazine, which paid them 1 million pounds for the exclusive rights.[47] Rival publication *Hello!* had unsuccessfully bid for the rights, offering the same amount. The couple agreed to engage a photographer themselves and supply *OK!* with pictures, and went to some lengths to comply with their obligation to use their best efforts to ensure that no other photographs would be taken. Despite this, a photographer infiltrated the wedding, who then sold the photographs he had taken to *Hello!,* which the latter published the following day.[48] A majority of the House of Lords confirmed that *OK!* also had a claim for breach of confidence against *Hello!*[49]

The distinction between legal protections for image rights and the breach of confidence claim is made clear in the following statement by Lord Hoffman:

> There is in my opinion no question of creating an "image right" or any other unorthodox form of intellectual property. The information in this case was capable of being protected, not because it concerned the Douglases' image any more than because it concerned their private life, but simply because it was information of commercial value over which the Douglases had sufficient control to enable them to impose an obligation of confidence.[50]

Even so, legal protections for confidential information protected "how the wedding looked—the photographic images which bring the event to life and make the viewer a virtual spectator at it."[51] This, in turn, vindicated the actors' concerns with controlling how they presented themselves to their public. While the specific legal claim was for breach of confidence, the following passage from the judgment explains some of the other interests at stake:

> Mr and Mrs Douglas sought to keep this information private primarily to protect their "image". Film directors take into account the public perception of actors and actresses

44. For a recent discussion of post-mortem image rights in the context of privacy principles, see Anita L. Allen & Jennifer E. Rothman, *Postmortem Privacy,* 123 MICH. L. REV. 285, 292–311 (2024). *See* CAL. CIV. CODE § 3344.1 (as amended by 2024 Cal. Stat. ch. 258 (A.B. 1836)) (West 2025).

45. NO FAKES Act, *supra* note 3.

46. OBG Ltd. v. Allan [2008] 1 AC 1.

47. *Id.* at 108.

48. In other proceedings, the Douglases recovered modest damages against *Hello! See* Douglas v. Hello! Ltd. [2007] UKHL 21.

49. For the related litigation initiated by Douglas and Zeta-Jones, see Douglas v. Hello! Ltd. (No 3) [2003] 3 All ER 996.

50. *OBG Ltd.*, 1 AC at [124].

51. *Id.* at [326].

> when casting for films. Miss Zeta-Jones said the "hard reality of the film industry is that preserving my image, particularly as a woman, is vital to my career". Mr Douglas said his name and likeness are valuable assets to him. It is important for him, for professional reasons, to protect his name and likeness and prevent unauthorised uses of either.[52]

As is true of defamation, however, the protections offered by a breach of confidence action do not entirely align with a right of publicity claim. The "image" of the Douglases was protected in a specific context: how they appeared at their wedding. No enforceable claim to confidence attaches merely to the appearance of world-famous actors. In this respect, *OBG Ltd. v. Allan* is thus another instance of gap-filling, in the absence of any direct protections for an individual's image.

The evolution of legal claims for breaches of confidence catalyzed the development of new tort protections for privacy rights.[53] In the leading decision from England and Wales, a newspaper was held to have invaded the privacy rights of celebrated fashion model Naomi Campbell when it disclosed that she was seeking treatment for drug use and published photographs of her outside the venue for a Narcotics Anonymous meeting.[54] The House of Lords held the newspaper article and the publication of the photographs to be unjustifiable intrusions into her private life.[55]

There are, however, significant limits to the capacity of privacy torts to protect an individual's image. The *Campbell* decision provides no support for the idea that disclosure of any image of the plaintiff may be prohibited: The newspaper was liable for breach of confidence only because the photograph added to the unjustifiable disclosure of private information about medical treatment.[56] The use of the image went beyond what could be justified by the news of Ms. Campbell seeking treatment for drug addiction.[57] In addition, rights to freedom of expression need to be balanced against interests in privacy, with some courts undertaking a proportionality analysis to balance the different public and private interests at stake.[58] And courts have reasoned that the disclosure must be accompanied by a reasonable expectation of privacy and, according to some lines of Commonwealth authority on privacy protections, the disclosure needs

---

52. *Id.* at [252].

53. In the United Kingdom, the Human Rights Act 1998 (U.K.) added to the impetus for the development of a new tort, as has the European Convention on Human Rights. *See* Campbell v. MGN Ltd. [2004] 2 AC 457, [11], [16], where Lord Nicholls of Birkenhead discussed the relationship between the evolution of the common law and jurisprudence under the Convention.

54. *Campbell*, 2 AC at [6]. The original newspaper story accompanying the photographs adopted a relatively supportive tone, directed at Ms. Campbell's determination to remedy her drug abuse. The newspaper published additional photographs and a much more critical story following Ms Campbell's filing of proceedings against it. *Id.* At 8.

55. Campbell had conceded that the newspaper was entitled to disclose that she was seeking treatment for drug dependency. *Campbell*, 2 AC at [129].

56. Relying on a first instance New Zealand decision, Lord Hope noted in *Campbell* that "[t]he taking of photographs in a public street must, as Randerson J said in *Hosking v. Runting* [2003] 3 NZLR 385, 415, [138], be taken to be one of the ordinary incidents of living in a free community." *Campbell*, 2 AC at [122].

57. This issue arose because the plaintiff had earlier publicly denied that she had resisted the prevalence of illegal drug-taking in the fashion industry. *Campbell*, 2 AC at [80].

58. *See* Nicole A. Moreham, *Privacy, Freedom of Expression and Legitimate Audience Interest*, 139 L.Q. REV. 412, 426-434 (2023).

to be highly offensive to a reasonable person in the position of the plaintiff.[59] As with defamation, the complex privacy law jurisprudence that has emerged in the Commonwealth jurisdictions will provide some protections for an individual's image in some, but certainly not all, circumstances.[60] And, again, to the extent that these developing common law principles can serve to protect an individual's image at all, the protections are only incidental to the vindication of privacy interests.[61]

C. PASSING OFF

An important milestone in the development of the capacity of the tort of passing off to protect image rights was the 1960 decision of the Full Court of the Supreme Court of New South Wales in *Radio Corp Pty Ltd. v. Henderson.*[62] The case concerned two well-known ballroom dancers whose image was reproduced without authorization on the sleeve of a record of ballroom dancing music. The plaintiffs succeeded on the basis that this would cause buyers of the record to believe that they "recommended the record as providing good music for ballroom dancing."[63] Referencing *Tolley v. Fry*, the court in *Henderson* noted the changing social and economic context since that decision:

> The point which seems to emerge with clarity is that one's conception of the status of an amateur sportsman 30 years ago is quite different to what is accepted today. The development in the advertising of products…has opened up a new field of gainful employment for many persons who…have found themselves in a position to earn

59. Hosking v. Runting [2005] 1 NZLR 1 at [42] (N.Z.). As Professor Nicole Moreham has observed, a reasonable expectation of privacy is not necessarily the same thing as a reasonable desire for privacy. Knowledge of media practices might mean that in some circumstances a reasonable person might not be able to expect any (or much) protection of privacy interests. *See* Nicole A. Moreham, *Recognising Privacy in England and New Zealand,* 63 CAMBRIDGE L.J. 555, 557 (2004).

60. New Zealand law also recognizes a tort of intrusion on a plaintiff's seclusion. In the leading authority, the actionable intrusion involved videoing a roommate while showering. C v. Holland [2012] 3 NZLR 672 at [1]. *See* Nicole A. Moreham, *A Conceptual Framework for the New Zealand Tort of Intrusion,* 47 VICT. UNIV. WELLINGTON L.R. 283, 283–84 (2016).

61. Privacy rights receive express protection under section 36 of the Québec Civil Code also, which designates use of a person's "name, image, likeness or voice for a purpose other than the legitimate information of the public" as an invasion of privacy. S.Q. 1991, c.3. The Supreme Court of Canada held in *Aubry v. Éditions Vice-Versa Inc* [1998] 1 S.C.R. 591 that the publication of a photograph of a seventeen-year-old child taken in a public place was an invasion of her privacy under the Code. The Court reasoned (at [53]):

> Since the right to one's image is included in the right to respect for one's private life, it is axiomatic that every person possesses a protected right to his or her image. This right arises when the subject is recognizable. There is, thus, an infringement of the person's right to his or her image, and therefore fault, as soon as the image is published without consent and enables the person to be identified.

The Court observed that it had to "be decided whether the public's right to information can justify dissemination of a photograph taken without authorization." *Id.* at [61]. The Supreme Court of Canada revisited the Aubry decision in *Syndicat Northcrest v. Amselem*, [2004] 2 S.C.R. 551 and elaborated further on the nature of the balancing test required. *See Id.* at [153]–[157].

62. Radio Corp Pty Ltd. v. Henderson [1960] NSWR 279, 292 (Austl.).

63. *Id.* at 281.

> substantial sums of money by lending their recommendation or sponsorship to an almost infinite variety of commodities.[64]

As with sporting celebrities, so with professional dancers: The activity of the plaintiffs had "resulted in their recommendation becoming a saleable commodity" that could be protected by the passing off tort.[65]

Australian courts continued the development of the passing off tort to protect a celebrity's image, most predominantly in cases involving the *Crocodile Dundee* franchise,[66] where passing off was deployed to prevent businesses suggesting that they were associated with the films. They had done so by deploying actors and imagery that imitated or invoked the actor who played the films' eponymous character.[67]

The capacity of passing off to protect image rights was further entrenched by a 2015 decision of the Court of Appeal of England and Wales[68] in a case involving unlicensed reproduction of a photograph of popstar Rihanna on t-shirts sold by a chain of clothing stores.[69] Dutifully reciting the conventional UK position, as set out in *OGB v. Allan*, the Court nevertheless confirmed that the use of a particular image of a celebrity could still give rise to a passing off claim, on the basis that consumers could think that the celebrity had authorized its use. Here, it was relevant that Rihanna was known to be a fashion icon; she had done promotional work for the clothing chain in the past; and customers would know that her styling in the particular photograph was reminiscent of images used in the context of her latest album. Accordingly, it was *because* Rihanna had engaged in promotional activities that a claim for passing off could be made out.[70] We are now a long way from *Tolley v. Fry.* Given the incessant commercialization of celebrity status by many famous individuals, in some circumstances there may be significant overlap between passing off and the right of publicity.

---

64. *Radio Corp*, [1960] NSWR at 292.

65. *Id.*

66. Pacific Dunlop Ltd. v. Hogan [1989] 23 FCR 553 (Austl.); Hogan v. Koala Dundee Pty Ltd. [1988] 20 FCR 314 (Austl.). On the overlap between common law remedies and Australian consumer protection laws, see Hutchence v. South Sea Bubble Co Pty Ltd. (1986) 64 ALR 330, 339–40 (Austl.). *See also* Andrew Terry, *Exploiting Celebrity: Character Merchandising and Unfair Trading,* 12 UNIV. N.S.W. L.J. 204, 212–17 (1989).

67. In one case, marsupial imagery, passing off was held to have occurred where a store used in its branding imagery of a koala dressed in a manner that invoked the eponymous character in the *Crocodile Dundee* movies. Hogan v. Koala Dundee Pty Ltd. [1988] 20 FCR 314 (Austl.).

68. An earlier first instance decision in the High Court of England and Wales (Chancery Division) to like effect was Irvine v. Talksport Ltd. [2003] EWCA (Civ) 423, [2002] 2 All ER 414 (passing off found where defendant had misrepresented that a prominent F1 racing driver was associated with a commercial radio station).

69. Fenty v. Arcadia Group Brands (trading as Topshop) [2015] EWCA (Civ) 3. The copyright in the photograph was owned by the photographer who then licensed it to the defendant.

70. The highest court in Malaysia, the Federal Court, has held that an individual celebrity has standing in a passing off claim of this nature, even if his celebrity status is exploited through a corporate vehicle. Mohammad Hafiz bin Hamidun v. Kamdar Sdn Berhad [2021] 4 MLI 878 [50]–[61] (Malay.). Citing passing off cases from other Commonwealth jurisdictions discussed in this Part, including *Irvine*, *Henderson,* and *Fenty*, the Court affirmed that in Malaysian common law the passing off tort could provide a remedy against misleading use of a celebrity's name or likeness—in this instance in the marking of fabrics.

Apex Commonwealth courts have yet to assess in detail any limits on the ability of the passing off that might be relevant in the deepfakes context.[71] Passing off remains, at least in theory, grounded in misappropriation of goodwill, which, on a conventional analysis, results from misrepresentations calculated to deceive or confuse the plaintiff's consumers. As noted above, the requirements of the passing off tort have been adapted to provide significant protections for a celebrity's image, but the tort has not sufficiently evolved to protect a celebrity's image *per se*. Issues may arise as to whether relief will be afforded where the use of a celebrity's image is accompanied by prominent disclaimers or in contexts (such as pornographic depictions) in which relatively few of those viewing the material would suppose that there was any connection between the plaintiff and the depiction at all. And, passing off's original concern with protecting business goodwill is likely to mean that the tort cannot provide any kind of remedy for "everyday" individuals lacking celebrity status.

Additional difficulties may arise from a UK Supreme Court decision holding that a passing off claim requires "customers" within the jurisdiction in which protection is sought.[72] On this view, one not shared by all Commonwealth jurisdictions, reputation is not enough.[73] It is not clear what this might mean for passing off claims by foreign celebrities. Celebrity seems to be squarely based on reputation, even if, as the Rihanna case shows, image rights are easier to protect when the plaintiff is in the business of endorsement, and had engaged in extensive promotional activity. In Rihanna's case, that activity had occurred in the United Kingdom. But consider the case of well-known Spanish or French movie stars who had *not* engaged in commercial activity within the jurisdiction. The UK Supreme Court's insistence on "customers" within the jurisdiction might preclude relief for passing off where an individual's fame derives from reputation alone.

***

In sum, in Commonwealth jurisdictions in which misappropriation of an individual's is not remedied by a separate form of tort liability, other torts—including defamation, breach of confidence, and passing off—provide some protections for misappropriation of an individual's image. But none of these causes of action, however much they have evolved, provides for direct protection of an individual's image. Individuals seeking to protect against misappropriation of their images are required to characterize their claims in line with conventional interests protected by these torts: reputation, confidences and privacy, and goodwill.

71. At the first instance level, there is some discussion of this point in a New Zealand decision, in which the judge cautioned: "Against creating any fresh monopolies in this area are freedom of expression, community access to intellectual progress and the public interest in competition." Tot Toys Ltd. v. Mitchell [1993] 1 NZLR 325 at 364 (N.Z.).

72. Starbucks (HK) Ltd. v. British Sky Broadcasting Group [2015] UKSC 31 [47].

73. *See* Graeme W. Austin, *The Consumer in Cross-Border Passing Off Cases*, 47 VICT. UNIV. WELLINGTON L. REV. 209, 211 (2016) (noting that the approach to territoriality in passing off cases in Australian and New Zealand case law differs from that of the UK Supreme Court).

II. MISAPPROPRIATION OF AN INDIVIDUAL'S IMAGE

Some Commonwealth jurisdictions do recognize misappropriation of personality as a distinct tort. In Canada, the common law has recognized a right of publicity from at least the early 1970s.[74] Summarizing the requirements of the tort, the Court of Appeal for British Columbia recently described the tort in terms that will be familiar to U.S. lawyers: "taking advantage of the plaintiff's name, reputation, likeness, or some other component of their individuality or personality that a viewer would associate with the plaintiff."[75] The earliest successful claim involved a professional water skier who had monetized his image, principally through promotion of specific photograph depicting his waterskiing skills.[76] Though the plaintiff was not widely known, the court held the defendants' use of a sketch based on the photograph to publicize its business to be an unlicensed appropriation of personality, a tort that was independent of passing off or infringement of copyright or trademark.[77]

A case involving the legendary Canadian pianist Glenn Gould provided an opportunity for the Ontario Court of Appeal to hold that commercial benefit alone is not a sufficient basis for liability, and distinguished between uses of a plaintiff's image for "sales" and uses where the plaintiff was the "subject" of the defendant's use.[78] Liability for appropriation of personality could arise only in the former category—where the defendant's commercial activity is not about the plaintiff, but instead uses the plaintiff's image to sell something else. Applying this framework, the court exonerated the defendant's publication of a book about Gould containing photographs of the plaintiff.[79]

In four Canadian provinces—British Columbia,[80] Saskatchewan,[81] Manitoba,[82] and Newfoundland and Labrador[83]—there are also statutory torts for the invasion of privacy. In the case of Saskatchewan, Manitoba, and Newfoundland and Labrador, the provincial statutes list the unauthorized appropriation of a person's name or likeness

74. Krouse v. Chrysler Canada Ltd. (1973) 1 O.R. 2d 225 (Can. Ont.). The claim in this case was ultimately unsuccessful, but the analysis provided the foundation for later cases. Useful discussions of the Canadian position include: Eric Reiter, *Personality and Patrimony: Comparative Perspectives on the Right to One's Image*, 76 TUL. L. REV. 673, 712–715 (2002); Robert Howell, *Personality Rights: A Canadian Perspective: Some Comparisons with Australia*, 1 AUST. INTELL. PROP. J. 212 (1990).

75. RateMDs Inc. v. Bleuer, 2025 BCCA 329, para. 121 (Can.).

76. Athans v. Canadian Adventure Camps Ltd. (1977), 17 O.R. 2d 425 (Can. Ont.).

77. The plaintiff does, however, need to be recognizable. *See* Joseph v. Daniels (1986), 4 B.C.L.R. 2d 239, para. 15 (no liability for use of image of the plaintiff bodybuilder's torso where the plaintiff could not be identified).

78. Gould Estate v. Stoddart Publ'g Co., 1996 CanLII 8209 (Can. Ont.) *aff'd* (1998), 39 O.R. 3d 545 (Can. Ont.).

79. In a recent case involving a website soliciting ratings of medical practitioners, the Court of Appeal for British Columbia considered that the website was on the "subject" side of the line. *RateMDs*, 2025 BCCA 329, para. 113.

80. Privacy Act, R.S.B.C. 1996, c 373, § 1(1).

81. Privacy Act, R.S.S. 1978, c P-24, § 2.

82. Privacy Act, C.C.S.M. c P125, § 2(1).

83. Privacy Act, R.S.N.L. 1990, c P-22, § 3(1).

as an example of an invasion of privacy.[84] The British Columbia privacy statute codifies the misappropriation of personality as a separate tort, creating a statutory tort where another uses:

> the name or portrait of another for the purpose of advertising or promoting the sale of, or other trading in, property or services, unless that other, or a person entitled to consent on the other's behalf, consents to the use for that purpose.[85]

Because a separate common law right of publicity has existed for several decades, the distinct role of statutory protections for image rights might be difficult to discern. But in recent case law questions have been raised about the preemptive effect of the statutory protections, implying that common law publicity rights should be tethered to, and possibly constrained by, the legislation.[86] The Court of Appeal for British Columbia has also noted, however, that the privacy interests protected by the British Columbia Privacy Act have a "quasi constitutional status and must be interpreted broadly,"[87] and, in this context, the court also noted the express and implicit privacy protections in the Canadian Charter of Rights, which are "informed by its underlying values of dignity, integrity, and autonomy."[88] At the same time, the link to the Canadian Charter necessitates taking account of the rights of others, including rights to freedom of expression and access to information.[89]

The Supreme Court of Appeals of South Africa has also confirmed that appropriation of personality gives rise to a distinct tort claim. In the leading case, *Grütter v. Lombard*,[90] the court's analysis drew on the common law and Roman Dutch tradition, as well as the protection of dignity in South Africa's Constitution.[91] The Court noted that in the new constitutional order, there are no sharp distinctions between privacy and preserving one's identity against unauthorized exploitation.[92] But the Court also observed that as with any civil wrong there may be circumstances in which considerations of public policy will justify conduct intrusion on a person's personality interests. None, however, applied on the facts of the case which involved straightforward appropriation of the plaintiff's name for personal advantage. *Grütter* was referenced in *Wells v. Atoll Media*, the decision discussed above involving the picture of the minor on the cover of the surfing magazine. In the latter decision, the court said:

---

84. Samuel Rowe, *Privacy in European, Civil, and Common Law*, *in* THE LAW OF PRIVACY AND THE MEDIA 72, 115 (Nicole A. Moreham & Adam Speker eds., 4th ed. 2025).

85. Privacy Act, R.S.B.C. 1996, c 373 § 3(2).

86. *See* Bao v. Welltrend United Consulting Inc., 2025 BCCA 3, [33]–[34].

87. *RateMDs*, 2025 BCCA 329, para. 36.

88. *Id.* at para. 53.

89. *Id.* at para. 98.

90. 2007 (4) SA 89 (SCA).

91. S. AFR. CONST., 1996 c 2 § 10.

92. *Grütter*, 2007 (4) SA 89 (SCA) at para. 12 (quoting Khumalo v. Holomisa, 2002 (5) SA (CC) at para. 27 (S. Afr.)).

> In the context of this case . . . the appropriation of a person's image or likeness for the commercial benefit or advantage of another may well call for legal intervention in order to protect the individual concerned.[93]

The common law of India also recognises personality rights.[94] A recent case involved misappropriation of the image and voice of Aishwarya Rai Bachchan, an extraordinarily famous Indian movie star.[95] Her claims were against the following parties: a website representing itself as Bachchan's official website; the host and distributor of unauthorized apps that enabled the downloading of wallpapers featuring Bachchan's image; a website selling t-shirts featuring photographs of Bachchan; an e-commerce platform selling and facilitating the sale of mugs and t-shirts featuring Bachchan's image; a motivational speaker firm using the plaintiff's name and image to suggest an association with Bachchan; a chatbox enabling users to engage with a digital impersonation of Bachchan, including in sexualized contexts; operators of a YouTube channel featuring deepfake videos depicting Bachchan, including "misleading and inappropriate content"; Google LLC in its capacity of owner of YouTube; and various John Doe defendants. Some of the allegations included complaints that defendants had superimposed Bachchan's face "on someone else's body, generating images of the Plaintiff with other celebrities and creating inappropriate content" and making sexually explicit comments and remarks.[96] Finding for the plaintiff, the court enjoined[97] misappropriation of the plaintiff's name/acronym, image and likeness and "any other attributes of her persona which are exclusively identifiable with her for any commercial and personal gain" including, but not limited to, artificial intelligence, generative artificial intelligence, machine learning, deepfakes, face morphing, in any medium and format.[98]

Judicial statements on the scope of the right of publicity tort in Indian cases reflect familiar concerns with the commercial value of celebrity. For instance, the court in the *Bachchan* observed:

> Personality rights . . . entail the right to control and protect the exploitation of one's image, name, likeness or other attributes of the individuals' personality, in addition to the commercial gains that can be derived from same.[99]

In this respect, the claim focused on the "commercial value of the picture or representation of a prominent person or performer and protects his proprietary

93. *Wells*, 2010 (4) All SA 548 (WCC) at para. 49 (S. Afr.) (noting the principle might not apply in crowd scenes).

94. A detailed survey of tort protections under the common law of India is provided in Karan Johar v. Indian Pride Advisory Pvt. Ltd., 2025 SCC OnLine Bom 546 (High Court of Judicature at Bombay, Ordinary Original Civil Jurisdiction, Commercial Division).

95. Aishwarya Rai Bachchan v. Aishwaryaworld, 2025 SCC OnLine Del 5943 (High Court of Delhi).

96. *Id.* ¶ 33.18.

97. This was on an interim basis. The procedural posture of the case was an application by the plaintiff for the court to waive the requirement of mandated mediation.

98. *Aishwarya Rai Bachchan*, 2025 SCC OnLine Del 5943, ¶ 39.

99. *Id.* at ¶ 34.

interest in the profitability of his public reputation or persona."[100] But Indian judges have also acknowledged the additional harms to the dignitary interests of the claimants, broadly in line with statements in Canadian and South African case law:

> When the identity of a famous personality is used without their consent or authorization, it may not only lead to commercial detriment . . . but also impact their right to live with dignity.[101]

The intertwining of commercial and dignitary concerns are also apparent in a 2023 case involving Anil Kapoor, another renowned Indian actor.[102] The case concerned misappropriation of his image in various contexts, including through the creation and dissemination of deepfakes. As in the *Bachchan* decision, the case targeted a variety of defendants who were alleged to have infringed the plaintiff's "personality rights, publicity rights and elements associated with his persona."[103] The plaintiff made broad claims as to his protectable interests: "his name; his voice; his photograph/image/likeness; his manner of speaking and dialogue delivery; his gestures; his signatures, etc."[104] The procedural posture of the case (an application for interim relief) did not provide an opportunity for the court to examine in detail the scope of image rights under the common law of India; at the same time, the court granted the relief sought by the plaintiff and made no suggestion that the actor's rights did not capture these aspects of his persona. The wide range activities for which relief was sought were similar to those detailed in *Bachchan*, and also included the making and dissemination of a variety of unlicenced digital replicas. [105]

---

100. Karan Jorhar, 2025 SCC OnLine Bom 546, ¶ 12 (citing DM Ent. Pvt. Ltd. v. Baby Gift House, CS (OS) No. 893 of 2002, ¶ 14).

101. *Aishwarya Rai Bachchan*, 2025 SCC OnLine Del 5943, ¶ 35.

102. Anil Kapoor v. Simply Life India, 2023 SCC OnLine Del 6914 (High Court of Delhi).

103. *Id.* ¶ 25.

104. *Id.*

105. The court provided the following inclusive list of the activities complained of by the plaintiff:

> Publishing and collecting fee by using his photographs that he would be attending an event, as a motivational speaker; Using morphed images of the Plaintiff and collecting monies for selling prints of his images; Creating wallpapers on mobile phones using Plaintiff's image; Using voice, dialogues and names from his movies in his own voice, as ringtones and ring back tones; Promoting and selling merchandise such as magnets, T-shirts, cups, stickers, keychains, using his photographs with/without the word 'Jhakaas'; Advertising and selling face masks with the Plaintiff's pictures; Providing electronic stickers with Plaintiff's image and collecting monies for the same; Using his name and photographs for posters; Selling suits under the Plaintiff's name and image; Providing forged autographs and photographs of the Plaintiff; Creating images and videos of the Plaintiff in a morphed manner; Using Artificial Intelligence to produce images and videos that are extremely derogatory, not merely to the Plaintiff but to other actresses as well . . . whose faces are being morphed with the Plaintiff's face - resultantly picturising the Plaintiff on a song or photograph with the clothes worn by these actresses; Generating images of the Plaintiff as cartoon characters using Artificial Intelligence; Providing GIF images of the Plaintiff on various social media handles; Squatting on domain names such as www.anilkapoor.in, www.anilkapoor.net and www.anilkapoor.com; Creating, publishing, and disseminating fake pornographic videos of the Plaintiff along with other actresses. *Id.* ¶ 30.

Drawing on earlier case law on the right of privacy in the Supreme Court of India,[106] the court in *Kapoor* linked the private law tort claim to the Constitutional guarantee of "life and liberty", from which the constitutional right to privacy is derived.[107] The tenor of the court's analysis is suggested by the following passage:

> The technological tools that are now freely available make it possible for any illegal and unauthorised user to use, produce or imitate any celebrity's persona, by using any tools including Artificial Intelligence. The celebrity enjoys the right of privacy, and does not wish that his or her image, voice, likeness is portrayed in a dark or grim manner, as portrayed on the porn websites. Moreover, the Plaintiff's image is being morphed along with other actresses in videos and images generated in a manner, which are not merely offensive or derogatory to the Plaintiff, but also to such other third party celebrities and actresses.
>
> The Court cannot turn a blind eye to such misuse of a personality's name and other elements of his persona. Dilution, tarnishment, blurring are all actionable torts which the Plaintiff would have to be protected against.[108]

Moreover, the protections afforded to the plaintiff were not just for him; they were also "for the sake of his family and friends who would not like to see his image, name and other elements being misused, especially for such tarnishing and negative use."[109] For present purposes, the following passage is perhaps the most significant: "The present case shows how elements of intellectual property that protect the attributes of an individual, in fact have other dimensions, including rights protected by the Constitution of India."[110] In line with the analysis in *Bachchan*, the court concluded: "If an injunction is not granted in the present case, it will lead to irreparable loss/harm to the Plaintiff and his family, not only financially but also with his right to live with dignity."[111]

III. CONCLUSION

Invoking the right to live with dignity in the context of personality rights reminds us that misappropriation of an individual's image or other attributes of personality is not only an economic concern. The dignitary concerns referenced in these cases might serve to expand our understanding of the conceptual foundations for protections of personality, while also exposing some of the common ground with the civil law approaches discussed by Professor Valérie Laure Benabou in this Symposium issue.[112] The defamation tort, which is sometimes relied on in jurisdictions in which publicity rights are not recognized, also serves to vindicate dignitary interest, as do emerging

106. R. Rajagopal v. State of Tamil Nadu, (1994) 6 SCC 632.
107. *Id.* ¶¶ 9, 26.
108. *Anil Kapoor,* 2023 SCC OnLine Del 6914, ¶¶ 42–43.
109. *Id.* ¶ 45.
110. *Id.* ¶ 46.
111. *Id.* ¶ 47.
112. Valérie-Laure Benabou, *Deep Fakes and Private Rights in the Perspective of EU Law: Is it Necessary to Intervene?*, 49 COLUM. J.L. & ARTS 729 (2026).

common law rights to privacy.[113] As David Louk's contribution to this Symposium underscores, countless "everyday" individuals, mostly women and girls, have been victimized by the deployment of deepfake technology.[114] As we get better at articulating the harms that deepfakes can inflict, legal responses will inevitably be required to take account of both the economic and dignitary interests at stake.

113. For an insightful analysis of the dignitary foundations of the common law privacy torts, *see* Nicole A. Moreham, *Why is Privacy Important? Privacy, Dignity and Development of the New Zealand Breach of Privacy Tort* in LAW, LIBERTY, LEGIS. (Jeremy Finn & Stephen Todd eds., 2008).

114. David S. Louk, Deepfakes, Real Enforcement Challenges, 49 Colum. J.L. & Arts 817 (2026).

Deepfakes and Transparency Obligations

Célia Zolynski*

ANNOTATED TRANSCRIPT

[ZOLYNSKI] First of all, I would like to thank Professor Jane Ginsburg, all the organizers, and the Paris Global Alliance program who made this comparative symposium possible. I'm delighted to be with you today. I propose to share my thoughts from an EU perspective about deepfakes and the current legal framework in the EU, and the potential evolution of this legal framework. This analysis is the result of current research I'm leading for the French Commission of Human Rights about teen intimacy and digital services,[1] a legal study for the French Agency for Food, Environmental and Occupational Health & Safety (ANSES) on the use of social media by adolescents and its health risks,[2] and another study for the Ministry of Culture about deepfakes in the creative sector.[3]

Let's begin with what we are talking about. From this morning's presentations we understand that the very notion of deepfakes is not so clear. There are many definitions of what deepfakes are and what the legal definition should cover. In this regard, the definition proposed by the AI Act is particularly interesting. In accordance with article 3(60) of the EU AI Act:[4]

---

* Professor of Law, Université Paris 1 Panthéon Sorbonne – Co-director of the Paris 1 AI Observatory.

1. CNCDH, *Avis sur la protection de l'intimité des jeunes en ligne*, A-2025-1 (January 2025), https://www.cncdh.fr/publications/avis-sur-la-protection-de-lintimite-des-jeunes-en-ligne-2025-1

2. Célia Zolynski et al., *Mineurs et Réseaux Sociaux: Étude des Dispositifs Légaux Relatifs à l'Usage des Réseaux Sociaux par les Mineurs*, IRJS [SORBONNE INST. LEGAL RSCH.] (Nov. 2025), https://www.anses.fr/system/files/Etude-juridique-IRJS-novembre-2025.pdf [https://web.archive.org/web/20260324230127/https://www.anses.fr/system/files/Etude-juridique-IRJS-novembre-2025.pdf].

3. *Le CSPLA Lance une Mission Relative aux Enjeux pour les Secteurs Culturels et Créatifs des Hypertrucages Générés ou Manipulés par l'intelligence Artificielle*, MINISTERE DE LA CULTURE (FR.), "(June 20, 2025), https://www.culture.gouv.fr/nous-connaitre/organisation-du-ministere/Conseil-superieur-de-la-propriete-litteraire-et-artistique-CSPLA/travaux-et-publications-du-cspla/missions-du-cspla/le-cspla-lance-une-mission-relative-aux-enjeux-pour-les-secteurs-culturels-et-creatifs-des-hypertrucages-generes-ou-manipules-par-l-intelligence-ar [https://perma.cc/2EAE-UTU2].

4. Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and amending Regulations 300/2008, 167/2013, 168/2013,

> 'deep fake' means AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful[.]

This is a very comprehensive definition focused on synthetic content.[5] It includes not only deepfakes of people, but any image, text, audio, or video created or manipulated by AI. At least two questions arise from this definition. First, where the text states "audio or video content that resembles existing persons," does that mean that a wholly invented person (a "synthetic person" as Professor [Jennifer] Rothman referenced) would not be a deepfake because it does not correspond to an actual person? The other objects of imitation in the definition, notably "events," however, suggest that the definition is intended to encompass the wholly fabricated as well as imitations based on actual "persons, objects, places, entities or events." To omit the wholly fabricated would mean that the text does not cover a great deal of false and misleading content. Such an omission seems inconsistent with the AI Act's overall goal to promote transparency to limit the risk of public manipulation.[6] Second, does the definition's statement that the content "would falsely appear" imply that a "deepfake" must have been created with a misleading purpose? Given the wording of article 3(60) of the AI Act, we can consider that this provision does not require proof that the producer of the content intended to mislead the public into believing that the content is real.

Beyond that, we need to ask whether deepfakes warrant regulatory attention. How are deepfakes novel or different from other kinds of false representations? Drawing an image or even creating a video does not present the same risk of being confused with reality because the representation is subjective. We understand that it is a representation perceived or constructed by the author. Therefore, the concept of representation brings us back to the classic debate about the relationship with the audience and fiction. Many consider that this debate is renewed with deepfakes, because some believe that hyper-realistic AI-generated or AI-manipulated content could make the public perceive the representation as real, or could prevent the public from taking a critical distance from the object. In other words, deepfakes could blur the line between fiction and reality.

Given this phenomenon, it seems important to take into account the context of what the deepfake was intended for, through understanding, as Jennifer Rothman explained this morning, that deepfakes are not a unique phenomenon. We therefore need to take into consideration the various purposes that deepfakes can serve. Several issues of concern should be mentioned here. A number of them appear to be particularly significant. One concerns the potential distortion of the information landscape, with the proliferation of synthetic media content that is now presented as if it were

---

2018/858, 2018/1139 and 2019/2144 and Directives 2014/90, 2016/797 and 2020/1828 [hereinafter "EU AI Act"], art. 3(60), 2024 O.J. (L 1689) 50.

5. Synthetic content refers to any form of media (text, images, audio, or video) that is either partially or fully created or significantly manipulated using artificial intelligence (AI) and machine learning techniques, rather than being captured directly from real-world events.

6. *See* Mateusz Labuz, *Deep Fakes and the Artificial Intelligence Act—An Important Signal or a Missed Opportunity?* 16 POL'Y & INTERNET 783 (2024); *infra* notes 10, 16 and accompanying text.

authentic; indeed, we observe the impact of a saturation of the digital space with the dissemination of a massive amount of inauthentic content which competes for viewers' attention in an already crowded online space. In addition, we need to take into account the massive infringement of individual rights that can result from the production and the sharing of non-consensual intimate images and child sexual abuse material.

Many of the issues surrounding deepfakes were beginning to be discussed years ago but are now more widely recognized. They are highlighted as a main concern by the International AI Safety Report 2025.[7] This report outlines a variety of risks to individuals and to society, such as misinformation, gender-based violence, and erosion of public trust in digital media, among others.[8]

The European authorities decided to tackle several of these issues by adopting the EU AI Act in 2024.[9] In this context, we observe that deepfakes are captured by the various layers of the AI Act (Part I) and that transparency has been considered a cornerstone of the regulation, whose effectiveness however remains limited, which leads to proposals for enhancements aimed at strengthening it (Part II).

I. DEEPFAKES CAPTURED BY THE VARIOUS LAYERS OF THE AI ACT

The AI Act is a regulation that aims to create a single EU market and harmonize rules to promote trustworthy and human-centric AI, based on compliance mechanisms and what we call a logic of risks. The goal is to promote innovation, but also to tackle the level of risks regarding safety and regarding human rights, democracy, and the rule of law. From this perspective, one of the principles of the AI Act is not to consider the technique itself, but its uses. Deepfake regulation is a perfect example of this regulatory approach of raising flags, sometimes red flags, but also of pushing innovation. However, the regulation of deepfakes illustrates how difficult it can be to promote both at the same time, especially considering that deepfake techniques can be used for various purposes and in various contexts.

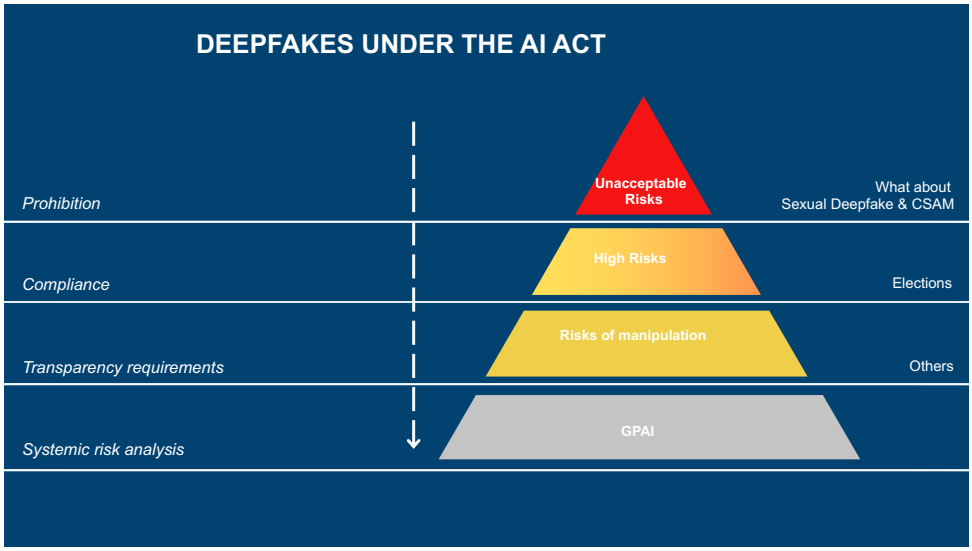
Sometimes deepfakes offer great opportunities that are socially desirable, for education, information, creation, and so on. But they can also cause massive harms, disinformation, fraud, bullying, harassment, and infringement to dignity. Because they

7. Yoshua Bengio et al., *International AI Safety Report,* AI ACTION SUMMIT (2025), https://internationalaisafetyreport.org/sites/default/files/2025-10/international_ai_safety_report_2025_english.pdf [https://web.archive.org/web/20260206184406/https://internationalaisafetyreport.org/sites/default/files/2025-10/international_ai_safety_report_2025_english.pdf].

8. Among harms to individuals caused by fake content, the report lists deepfakes' propensity to: extort, scam, psychologically manipulate, or sabotage targeted individuals or organizations; impersonate authority figures or trusted individuals to commit financial fraud; blackmail individuals for extortion purposes; sabotage individuals in their personal and professional lives, violating rights 'to one's honor and reputation; create abuse through fake pornographic or intimate content overwhelmingly targeting women; and distinctly harm children via AI-generated sexual content. *Id.* at 62–63.

9. EU AI Act, *supra* note 4.

can have non-harmful purposes, deepfakes have not been prohibited per se by the AI Act, even though it was suggested. Nevertheless, the EU authorities have identified the need to preserve the public interest by avoiding specific risks of manipulation and malicious uses. This was one of the main goals of the AI Act, both during its negotiation and at the moment of its adoption in June 2024. The idea was to divide AI into risk levels and to prohibit, among other things, those uses that cause significant harm through manipulation, impersonation, and deception. There was also great concern about the integrity of the information ecosystem and of elections. Considering these risks, deepfakes could be captured by the different layers of the AI Act regulation.[10]



The AI Act recognizes that some deepfakes could be considered in the category of ""high risk,"" a qualification that determines the application of most of the compliance requirements implemented by the regulation. In doing so, the EU legislature recognizes that deepfake technology is not problematic in itself and that deepfakes are not a unique phenomenon. However, certain contexts of use lead us to consider the risk that requires regulation. This is a perfect illustration of the AI Act regulation logic that takes into account the context of the use. The use of deepfakes in an electoral context specifically falls under this category due to their potential for manipulation and the spread of electoral disinformation (considering the dangers to the integrity of votes and democratic processes that we could face). For all other uses, the AI Act imposes transparency requirements to address the risk of manipulating the public exposed to deepfakes.

This raises the question of whether the use of certain types of content should be prohibited. In this regard, it should be noted that the AI Act defines a list of prohibited uses of AI, by identifying so called "unacceptable risks" regarding safety and

10. For critiques of such classifications, see Labuz, *supra* note 6.

fundamental rights. Those are set out in article 5 of the AI Act, which includes no mention of deepfakes.[11] This omission is very controversial and may be revisited considering harmful issues of non-consensual sexual deepfakes and child sexual abuse material (CSAM) [12]. In February 2025, the EU Commission published guidelines to interpret article 5 of the AI Act that mention NCII (Non-Consensual Intimate Imagery) and CSAM as possible prohibited uses.[13] But the conditions necessary to apply the prohibitions of article 5 are very strict, and we are not sure that NCII and CSAM could satisfy all. In considering this, we must promote the evolution of the text to address such harmful uses.

Finally, this whole risk assessment structure is completed by imposing a specific obligation on providers of GPAI (General-Purpose AI) systems and GPAI models that could cause systemic risks,[14] now described by the Safety and Security Chapter of the Code of Practice published by the EU in July 2025.[15] The Code mentions that specific uses of deepfakes such as CSAM and NCII can generate systemic risks. This recognition arrives late, but makes it possible to take such deepfakes into account indirectly.

II. TRANSPARENCY: A CORNERSTONE OF REGULATION TO CAPTURE DEEPFAKE RISKS

Under the AI Act, it is therefore important to note that most deepfakes are regulated only by the Act's transparency requirements. The goal is to ensure trustworthiness of AI systems by maximizing transparency as users are exposed to their outputs. This is in

11. EU AI Act, *supra* note 4, art. 5.

12. This will be covered by EU Member States legislation. *See* Directive (EU) 2024/1385 of the European Parliament and of the Council of 14 May 2024 on Combating Violence Against Women and Domestic Violence, O.J. (L 1385). For instance, in France, the publication of synthetic intimate imagery (NCII) has been a criminal offense since 2024. CODE PÉNAL, art. 226-8-1 (Fr.). *See* Zolynski et al., *supra* note 2 and CNCDH, *supra* note 1.

13. *See* EUROPEAN COMM'N, *Guidelines on the Definition of an Artificial Intelligence System* (2025), https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application [https://web.archive.org/web/20260219200600/https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application]; EUROPEAN COMM'N, *Guidelines on Prohibited Artificial Intelligence Practices Under Regulation (EU) 2024/1689* (Feb. 2025), https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act [https://web.archive.org/web/20260219200600/https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act].

14. The AI Act describes "systemic risks": "General-purpose AI models could pose systemic risks which include, but are not limited to, any actual or reasonably foreseeable negative effects in relation to major accidents, disruptions of critical sectors and serious consequences to public health and safety; any actual or reasonably foreseeable negative effects on democratic processes, public and economic security; the dissemination of illegal, false, or discriminatory content." *See* EU AI Act, *supra* note 4, recital 110.

15. EUROPEAN COMM'N, *Code of Practice on General Purpose AI* (July 2025), https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai [https://web.archive.org/web/20260219185012/https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai]. Under the EU AI Act, they will have to identify and mitigate those risks. *See* Regulation (EU) 2024/1689, art. 55.

order to protect the public from malicious users and disinformation.[16] We need to understand here whether these transparency requirements should be the cornerstone of the regulation. In other words, does such transparency offer sufficient remedies, or should it be considered insufficient to ensure the interests of the public?

My first aim is to identify the questions raised by the AI Act's transparency approach imposing specific requirements to address deepfake risks. My second aim is to determine how transparency could be an effective means of addressing the potential risks we have mentioned, by considering not only obligations of technical transparency, but also the ensemble of governance mechanisms ensuring public access to information through external, independent oversight to build a so-called "*regulation through transparency.*"

A. TRANSPARENCY REQUIREMENTS (HOW AND WHAT FOR?)

Article 50 of the AI Act introduced two levels of transparency requirements that will take effect in August 2026.[17] First, AI providers must design their AI systems, including general purpose ones, to mark, in a machine-readable format that can be detected, outputs as artificially-generated or manipulated.[18] The lengthy AI Act Recitals specify what kind of technical tools can be used, using techniques such as watermarking, for example.[19] The aim here is to facilitate trustworthy detection and identification of AI-generated and manipulated content. In addition, there is a labeling requirement imposed on the deployers of AI systems.[20] They must label deepfakes in such a way that the public can be informed of the synthetic nature of the content. Under article 50(5), information must be provided in a clear and distinguishable manner at the latest at the time of the first interaction of exposure.[21]

Because these provisions could be quite difficult to implement for these actors, guidelines and a code of conduct will be published to better define such requirements, helping deployers and providers of GenAI systems to detect and label AI or manipulated content.[22] These texts, which are considered soft law in European Union law, will

16. Labuz, *supra* note 6, at 791 ("The very basic idea of transparency obligations in relation to deep fakes is to enable the recipients to make informed choices on displaying the material and spot that the epistemic value of the material displayed has changed. It is therefore expected that appropriate markings will serve as a warning to users and a safeguard against dis- and misinformation.").

17. EU AI Act, *supra* note 4, art. 50.

18. *Id.* art. 50(2).

19. *Id.* recital 133.

20. *Id.* art 50(4) (defining a deployer as "a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity").

21. *Id.* art. 50(5).

22. EUROPEAN COMM'N, *FAQs: Guidelines and Code of Practice on Transparent AI Systems* (2025), https://digital-strategy.ec.europa.eu/en/faqs/guidelines-and-code-practice-transparent-ai-systems [https://web.archive.org/web/20260210150922/https://digital-strategy.ec.europa.eu/en/faqs/guidelines-and-code-practice-transparent-ai-systems]; EUROPEAN COMM'N, *Stakeholder Consultation on Transparency Requirements for Certain AI Systems under Article 50 of the AI Act* (2025), https://digital-strategy.ec.europa.eu/en/news/commission-launches-consultation-develop-guidelines-and-code-practice-

specify the technical requirements and how to take into account the context of the deepfake. This seems essential considering the various challenges we already can identify with respect to the transparency requirements imposed by the AI Act.

In particular, many limits and challenges have been identified respecting the effectiveness of these new requirements imposed by the AI Act considering that the goal pursued here, as mentioned, is to ensure the trust of the user and the general public and to avoid the risk of manipulation.

Some concerns have already been taken into account by the AI Act itself. For example, the text addresses the need to preserve freedom of expression, freedom of creation, and freedom of science.[23] In this context, the AI Act still requires labeling, but specifies that the label can be adapted so as not to hamper the display or enjoyment of the work.[24] Thus, this is not a total exemption in case of creative or scientific content. This approach prompts the question: What is artistic content? The AI Act provides that the content has to be a work that is *evidently* artistic, creative, satirical, fictional, or analogous. While "evidently" suggests an endeavor to avoid pretextual claims of artistic content, the open-ended phrase "or analogous work" risks expanding the universe of content benefitting from a relaxed labelling requirement.

There are also other limits and challenges to transparency, such as technical ones. The Coalition for Content, Provenance, and Authenticity (C2PA) is currently working on watermarking.[25] In France, for example, the Provenance for Trust initiative aims at proposing to create a coalition of researchers and authentication services including the Journalism Trust Initiative (founded by Reporters without Borders),[26] working with experts on labeling content and detecting AI-generated content, to propose open source technical solutions and specific certification mechanisms for media.

There are a lot of technical challenges to consider, especially regarding the robustness and accuracy of watermarks.[27] More generally, we need to take into account other challenges to reach the goal pursued, such as the limits of label's ability to inform

transparent-ai-systems [https://web.archive.org/web/20260219152742/https://digital-strategy.ec.europa.eu/en/news/commission-launches-consultation-develop-guidelines-and-code-practice-transparent-ai-systems].

23. For example, deepfakes can be used in video games to create a non-player character using motion picture techniques. In a historical context, deepfakes can be used to produce counterfactual history. *See, e.g.*, LE MONDE SELON L'IA: EXPLORER LES ESPACES LATENTS [THE WORLD THROUGH AI: EXPLORING LATENT SPACES] (Jeu de Paume & JBE Books, 2025).

24. *See* EU AI Act, *supra* note 4, art. 50(4) ("Where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the transparency obligations set out in this paragraph are limited to disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work.").

25. COAL. FOR CONTENT, PROVENANCE & AUTHENTICITY, https://c2pa.org/ [https://web.archive.org/web/20260221005527/https://c2pa.org/] (last visited Mar. 24, 2026).

26. PROVENANCE FOR TRUST INITIATIVE, https://www.provenance4trust.org/ [https://web.archive.org/web/20260128180453/https://www.provenance4trust.org/] (last visited Mar. 24, 2026).

27. EUROPEAN PARLIAMENT, *Generative AI and Watermarking* (2023), https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf [https://web.archive.org/web/20260204172313/https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf]; EUROPEAN COMM'N, *Stakeholder Consultation*, *supra* note 22.

the public of the nature of the deepfake. This means that we need specific analysis to ensure that the goal pursued can be reached, whereas Article 50 of the AI Act is very general concerning the information to be provided.[28] Here, academics and researchers are of critical importance in identifying appropriate labels and addressing cognitive biases.[29] We also have to address the enforcement problem considering that "malicious actors will simply not adhere to the obligations."[30] Beyond that, it is essential to consider an epistemic issue regarding the impact on public opinion of the dissemination of a massive amount of fake synthetic content even if the deepfakes are labeled as such. In this context, transparency requirements might not be enough to prevent public manipulation.

That is the reason why it is critical to take action for the education and resilience of the public and to consider the relationship between users and the information space, in order to ensure better user agency.[31] Regarding all of this, we need to go a step further to understand if transparency is a sufficient remedy to limit the harmful impact of deepfakes. In order to reach this goal, we need to assess transparency not only as meeting specific technical or legal requirements but also whether it achieves genuine transparency in fact.

B. REGULATION THROUGH TRANSPARENCY (TAKE IT A STEP FURTHER)

Considering these limits and challenges, we need to go a step further to take the public interest into account by promoting not only transparency requirements but also regulation through transparency.

First and foremost, it means involving third parties (regulators, academics, NGOs) in reviewing the transparency measures imposed by the AI Act to be adopted in order to achieve the intended objective—namely, to avoid the risk of public manipulation—and to ensure their effectiveness.

To this end, AI systems and model providers should first publish periodic transparency reports detailing the various measures taken and why they were chosen. This should be supplemented by the periodic publication of risk assessment reports required for providers subject to the AI Act. These reports would describe how providers have identified the systemic risks caused by the generation of deepfakes from their models and what mitigation measures have been taken to effectively limit these

28. Martina J. Block, *A Critical Evaluation of Deepfake Regulation Through the AI Act in the European Union*, 4 EUROPEAN CRIM. L. REV. 184 (2024).

29. EUROPEAN COMM'N, *Guidelines for Providers of VLOPs and VLOSEs on the Mitigation of Systemic Risks for Electoral Processes* (Apr. 2024), https://digital-strategy.ec.europa.eu/en/library/guidelines-providers-vlops-and-vloses-mitigation-systemic-risks-electoral-processes [https://web.archive.org/web/20260220131404/https://digital-strategy.ec.europa.eu/en/library/guidelines-providers-vlops-and-vloses-mitigation-systemic-risks-electoral-processes].

30. Block, *supra* note 28.

31. *User Empowerment through Media and Information Literacy Responses to the Evolution of Generative Artificial Intelligence (GAI)*, UNESCO (2024), https://unesdoc.unesco.org/ark:/48223/pf0000388547.locale=en.

risks. However, for these reports to be fully effective, the regulator must develop a common methodology for producing them in a way that is fully useful. In addition, access to data should be ensured not only to regulators but also to academics and non-profit organizations in order to evaluate how responsibly these online service providers have been in implementing their transparency obligations.

Furthermore, this regulatory approach based on transparency must extend beyond the AI Act, as the production and dissemination of deepfakes require the development of a systemic regulatory framework in light of massive sharing on online services, especially on social media. In this perspective, beyond the AI Act, it is essential to take into account the new requirements imposed by the Digital Services Act ("DSA") on online platforms that aim at ensuring the safety of the digital space, in particular to limit the impact of deepfakes' virality and algorithmic amplification. To that end, first of all, very large online platforms have to undertake risk analysis and take mitigation measures to limit the systemic risks produced by the use of their services such as threats to privacy, dignity, media pluralism, and health.[32] Such obligations have been set out in the specific context of elections to consider the risks of massive amount of deepfakes for the integrity of election results and democratic debate.[33] The Guidelines of article 28 of the DSA,[34] imposing specific obligations to online platforms accessible to underage end-users, also cover deepfakes.[35] This is especially essential considering that "children are particularly vulnerable to synthetic content, such as deepfakes, which can make them more exposed to harmful online practices like grooming and cyberbullying, as well as child sexual abuse material (CSAM)."[36]

In considering deepfake-related legislation, it is particularly important to monitor the effectiveness of mitigation measures online service providers take to address these risks, for example, for individuals or for democratic debate. In an effort to achieve these goals, the DSA also imposes external and independent audits. Such audits are essential notably to assess the effectiveness of the guardrails implemented by the AI provider, for example, to avoid specific kinds of deepfakes.[37] This is one of the major components of the regulation through transparency by evolving various stakeholders. This appears

32. For example, they must label deepfakes and ensure that this label stays with the content even if the content is shared with other users. *See* Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services (hereinafter "Digital Services Act"), arts. 34–35, O.J. (L 277).

33. EUROPEAN COMM'N, *Guidelines for Providers*, *supra* note 29.

34. Digital Services Act, *supra* note 32, art. 28.

35. EUROPEAN COMM'N, *Guidelines on Measures to Ensure a High Level of Privacy, Safety and Security for Minors Online* (July 2025), https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:C_202505519 [https://perma.cc/9SVZ-99C7].

36. EUROPEAN PARLIAMENT, *Children and Deepfakes* (July 2025), https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2025)775855 [https://web.archive.org/web/20251211125034/https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2025)775855].

37. *See, e.g.*, CTR. FOR COUNTERING DIGITAL HATE, *Fake Image Factories* (2024), https://counterhate.com/research/fake-image-factories/ [https://web.archive.org/web/20251229085406/https://counterhate.com/research/fake-image-factories/].

to be a decisive step toward protecting the public interest by preventing manipulation and disinformation.

Transparency as a Regulatory Duty

Olivier Sylvain*

ANNOTATED TRANSCRIPT

[SYLVAIN] Hi, everyone. It's great to be here, and I'm so pleased to have been invited to join this conversation. I come to you not as an IP specialist; I am a public law minded scholar, and I tend to think of most policy problems as public law problems.

In this regard, I'm attending to the problem of deepfakes and related AI abuses as a problem of deception on the public. Jennifer [Rothman] in the first panel asked if deepfake harms are limited to just individual victims. This is a great way to frame the issue. I recommend that, to the extent we're thinking about people, we ought to think about harms to the greater public.

The point I want to emphasize here is that there are very few institutions that are better able to attend to public consumer harms as a matter of course than federal and state government agencies. In this regard, they have the authority to mandate and scrutinize companies' public disclosures and risk assessments. More than this, they are best situated to stand in the shoes of or act on behalf of consumers who otherwise generally do not have the time, expertise, or wherewithal to attend to the workings of companies.

One challenge here is that agency officials are only as effective as their leaders choose to be. The current presidential administration's direct assault on the administrative state and the civil servants who make it work make this concern plain as day.

But there's also another formidable challenge, and that's what I'm going to take up here: To what extent does the First Amendment allow federal agencies to regulate the kinds of information that we've been talking about today?

In order to really put a context to all this, I want to make clear that I'm not coming to the problem of deepfakes and related harms as an IP problem as most of you are. I say this because transparency measures like mandated disclosures or risk assessments do not mean to provide regulatory benefits apart from the information they convey. They have salutary behavioral impacts. Risk assessments, for example, do more than

* Professor of Law, Fordham University School of Law; Senior Research Fellow, Knight First Amendment Institute at Columbia University.

report on risks; they arguably nudge companies to attend to potential harms. This is not a new revelation. In environmental regulation, the National Environmental Policy Act of 1970 sets out the impact assessment obligation on the same theory.[1] About the same can be said for disparate impact assessments and privacy assessments. These are all designed not merely for the substance of the disclosure to the public, but also presumably to be habit-forming.

Now, I will offer a simple taxonomy of transparency. Here, I will build on ideas from Celia [Zolynski]'s presentation.

One is *mandated disclosures*, which I think risk assessments could fall under. But there are many other disclosures, nutrition labels and breach disclosures, for those of you who attend to cybersecurity issues.

Next is *audit requirements*: They are a kind of disclosure, but they don't do the same thing as a mandated risk assessment. These might be undertaken internally or by a third-party for publication to government officials or the general public.

There's also *appellate process*. Any given platform or company that decides to take something down ostensibly ought to be able to give individuals whose content is taken down the opportunity to appeal after some explanation. The Digital Millennium Copyright Act, as many of you likely know, has a mechanism to publicize, although to just one person, a potential aggrieved party, the possibility of a takedown.[2] Counternotification is something that comes up in the Take It Down Act.[3] This is a form of transparency, even if it sounds in due process. Danielle Citron has written about administrative due process in this context.[4]

Another *is civil investigative demands*. Some civil law enforcement agencies like the Federal Trade Commission (FTC) have the authority to scrutinize companies through civil investigative demands. They issue these, on the one hand, to determine whether to commence a formal enforcement action, as well as to produce a report about industry practices to Congress and the greater public.

And finally, *data access*. I often think of this as related to researchers—that researchers have access to the ways in which platforms use data. There's a lot of learning that has to happen in that space. I'm a senior fellow at the Knight First Amendment Institute here at Columbia, and that is one of the priorities for them, for example.

Given this taxonomy, my focus here is going to be very narrow, and it's going to be on the kinds of mandated disclosures and risk assessments that we've already been hearing about. And it's going to be also narrowed in the context of threats to elections and consumer harm. I also am not talking here about provenance for the purposes of IP holders or creators' rights. I'm really instead talking about public harms—the kinds of harms for which agencies and governments ostensibly stand in the shoes of consumers. Disclosures have social benefits that are unrelated to the specific harm to creators and other IP rights holders.

1. National Environmental Policy Act of 1969 § 106(b)(2), 42 U.S.C. § 4336(b)(2).
2. 17 U.S.C. § 512(g).
3. Take It Down Act, S. Res. 146, 119th Cong. (2025) [hereinafter "Take It Down Act"] (enacted).
4. *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1249–50 (2008).

In this regard, one important difference between the EU and the US, of course, is the First Amendment. This is why transparency is a tricky regulatory intervention here. I'm going to start by talking about this in the context of social media regulation, because that's actually the area in which the Supreme Court has recently given us some guidance about what transparency requirements may or may not do.

As many of you likely know, the big case that the Supreme Court decided in the summer of 2024 involved state legislation out of Texas and Florida on content moderation practices of the big platforms.[5] Now, the regulations prohibited companies from blocking posts, taking down content, or suspending users based on the viewpoints the users expressed.[6] This is mainly what Justice Kagan's opinion is addressed to.[7] But those statutes also had transparency provisions. They required social media platforms to provide users with notice and individualized explanation for why content would be taken down.[8] Texas's law also required platforms to afford users the opportunity to appeal those decisions.[9] We have some language in the Supreme Court opinion about this. Not a lot.

Now NetChoice, industry folks, and First Amendment advocates brought cases against the state laws, arguing that they violated the First Amendment because they burdened editorial decisions of the companies.[10] A company that has to explain every takedown decision carries a heavy burden that could chill their editorial decision making. This is actually pretty intuitive in First Amendment doctrine. The *Zauderer* case is the principal case to talk about mandated disclosures as a species of commercial speech regulation.[11] The *Zauderer* test requires a kind of balancing given the speech interests at stake.

The Eleventh Circuit, reviewing the Florida case, said that the individualized explanation requirements were unduly burdensome. The Fifth Circuit did not think so. They did not think that Texas's approach to this was burdensome.[12]

Writing for the Supreme Court, which was unanimous in its judgment, Justice Kagan puts a lot of cold water on any effort to regulate social media news feed content moderation. That is the main takeaway of the decision. The Court remanded the case because the challenge that NetChoice and others brought was too broadly addressed to a wide range of editorializing, including news feed moderation.[13] To do a proper facial challenge analysis, the courts below have to determine whether a substantial range of applications would be affected by this regulation. And that is where we are now.

---

5. Moody v. NetChoice, LLC, 603 U.S. 707, 707 (2024).
6. *Id.* at 720–21.
7. *See id.* at 726–43.
8. *Id.* at 720–21 (citing FLA. STAT. §§ 501.2041(2)(d), 501.2041(3) (2023) and TEX. BUS. & COM. CODE ANN. § 120.103(a)(1)).
9. *Moody,* 603 U.S. at 721 (citing TEX. BUS. & COM. CODE ANN. §§ 120.103(a)(2), 120.104).
10. *Moody,* 603 U.S. at 721.
11. Zauderer v. Office of Disciplinary Counsel of Sup. Ct. of Ohio, 471 U.S. 626 (1985).
12. *Moody,* 603 U.S. at 722.
13. *Id.* at 744.

But part of the analysis was the consideration of whether or not the disclosure requirements or the explanation requirement imposed a burden on the speech interests of the companies.[14] We don't have an answer from Justice Kagan's opinion, but we have strong indications that the Supreme Court would strike down the transparency requirements because of the burdens it imposes on companies. Justice Thomas, who is apt to invite all kinds of litigation involving matters that worry him, has said that he would want to revisit the *Zauderer* test for how to evaluate whether something is too unduly burdensome of commercial speech.[15] And he's skeptical that *Zauderer* actually articulates a view that is consistent with First Amendment norms.[16] He would actually, if not do away with it, substantially narrow the claim that there's a burden on speech interests, which is an interesting intervention.

So the split here between the Eleventh and Fifth Circuit is actually a story that reveals tension in cases across the states.

In this regard, I will shift to California's transparency laws on deepfakes. But it's worth saying that twenty-six states have passed laws regulating political deepfakes in particular.[17] Many of these have prohibit deepfakes in elections, and, more pertinently, contain disclosure and transparency requirements.

Congress, too, has been thinking about this. The Protect Elections from Deceptive AI Act is a bipartisan bill that would prohibit the distribution of materially deceptive media that is generated by AI relating to federal candidates.[18] The federal candidate can bring an action, which brings up all sorts of things that came up before about how to vindicate harms. And there's a First Amendment exception for parody and content involving news broadcasts.[19]

As to California, there are two statutes that were passed late last year: A.B. 2839 and A.B. 2655.[20] AB 2655 is the Defending Democracy and Deepfake Deception Act, or "DDDA." It requires large platforms to label certain content as inauthentic, fake, or false, during the 120 days of before an election, and disclosure requirements after the election.[21] The content that portrays candidates for elective office and current elected officials has to include a statement that says "this image, audio, or video has been manipulated and is not authentic."[22] Given what you've heard from me about burdens on speech, it is suggestive that this is potentially the kind of thing that the doctrine wouldn't allow.

---

14. *Id.* at 725–26.

15. *Id.* at 751 (Thomas, J., concurring).

16. *Id.*

17. *See Tracker: State Legislation on Deepfakes in Elections*, PUB. CITIZEN (updated Mar. 10, 2026), https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/ [https://perma.cc/KY9Y-QEAW].

18. Protect Elections from Deceptive AI Act, S. Res. 1213, 119th Cong. (2025); H.R. Res. 5272, 119th Cong. (2025).

19. *Id.*

20. A.B. 2839, 2023–2024 Reg. Sess. (Cal. 2024); A.B. 2655, 2023–2024 Reg. Sess. (Cal. 2024).

21. CAL. ELEC. CODE § 20513(e) (West 2024).

22. ELEC. § 20513(d).

Indeed, this is something that has been the subject of lawsuits. The California law also has a substantive deceptive media and advertisements provision.[23] I think maybe the next panel might mention a bit about that later. I don't need to say much here on that other than that this law, too, has a transparency provision. Federal courts have found it to be unconstitutional as a matter of First Amendment doctrine because it is viewpoint-based and content-based.[24]

With regards to the transparency provisions, there's a weird order from the bench from Judge Mendez, the same judge, that says that the legal challenge can't move forward because of section 47 USC § 230, a provision that Jennifer [Rothman] mentioned and about which I have written quite a bit.[25] It preempts the state's effort to regulate the distribution of user generated content.[26]

On the Take It Down Act, we have cases that are addressed to transparency. We have a standard for evaluating whether it's unduly burdensome for speakers. We don't have any clear direction from the Supreme Court. But we have an inkling, given *Moody v. NetChoice*––that is, the case involving the Texas and Florida social media laws. The Tools to Address Known Exploitation by Mobilization Technological Deepfakes on Websites, and Networks Act—that's the full title of the Take It Down Act—criminalizes the non-consensual distribution of intimate images, whether authentic or digitally manipulated.[27] There are definitions of what is an intimate visual depiction that's drawn from another provision in the U.S. Code.[28] The Consolidated Appropriations Act has a definition of this. And there are distinctions in the statute between intimate visual depictions of adults and those involving children with regards to adults.[29]

Among other things, the intimate visual depiction has to have been obtained or created under circumstances in which the person who posted it knew or reasonably should have known that the identifiable individual had a reasonable expectation of privacy.[30] So the person, whoever posted it, had some understanding that the other person had an expectation of privacy. Alternatively, the law provides that the inauthentic intimate visual depiction disclosed is without consent.[31] An authorization mechanism.

With regards to minors, the statute says that knowingly publishing intimate visual depictions with the intent to abuse, humiliate, harass, or degrade the minor, or arouse

23. ELEC. §§ 20012(b), (d).

24. *See* Kohls v. Bonta, 797 F. Supp. 3d 1177 (E.D. Cal. 2025).

25. See, e.g., Olivier Sylvain, Reclaiming the Internet: How Big Tech Took Control—and How We Can Take It Back (2026); Olivier Sylvain, Platform Realism, Information Inequality, and Section 230 Reform, 131 Yale L. J. FORUM 475 (2021); Olivier Sylvain, *Intermediary Design Duties*, 50 CONN. L. REV. 203 (2018).

26. Order and Final J. and Permanent Inj. as to AB 2655, Dkt. 98, Kohls v. Bonta, No. 2:24-CV-02527-JAM-CKD, 2025 WL 2495613 (E.D. Cal. Aug. 20, 2025).

27. Take It Down Act, *supra* note 3.

28. *Id.* § 2(h)(1)(E).

29. *Id.* § 2(h)(2).

30. *Id.* § 2(h)(2)(A)(i).

31. *Id.* § 2(h)(3)(A)(i).

or gratify the sexual desire of any person is a violation.[32] And, for what it's worth, this is consonant with other ways in which, I think, the public laws address harms to children and obscenity laws more generally. There's a fine, a criminal fine, and criminal imprisonment as a possibility.[33] The civil penalty, I'm not completely sure about, but the offenses involving adults can put someone in prison for no more than two years, and those involving minors no more than three years. That's the substantive obligation.

Now I'll turn briefly to notice and takedown. The Take It Down Act, which I should have said, was passed with the President's signature in April to great fanfare.[34] The part of the law that has gotten the most attention are the provisions that require cover platforms to remove non-consensual, intimate visual depictions within forty-eight hours of having notice of it.[35] The difference between these notice and takedown provisions and the criminal provisions is that, in the former, there is no similar cabining of what an intimate image for the purpose of the statute is. And this is going to be important for thinking about the vulnerabilities of this law.

The law also requires companies to pose clear and conspicuous information about the removal process. The FTC has enforcement authority to issue penalties for non-compliance. There is no private right of action. There is a safe harbor for platforms that, in good faith, remove content when they have notice of it: This parallels the so-called immunity under section 230 for interactive computer services.[36] By the way, this provision is an amendment to section 223, which is the neighbor of section 230, for those of you who pay attention. And the last thing I'll say about this is that the law passed—remarkably—with the support of a bipartisan consensus.

You might accordingly think that this would mean everything is in the clear. After all, everybody wants to protect the kids.

But there are some flaws here, and I'll just identify a couple. The companies did not love the forty-eight-hour takedown requirement. Once you have notice, you have forty-eight hours to take it down.

There is a potential overbreadth problem here given that the criminal provisions cover a narrower scope of activity than the notice and takedown provisions do. And so, you might see protected content getting taken down. Consider the example of a journalist's photograph of a topless protester.

Now, there is a more pernicious problem, and it is what makes this upside down in many regards for people who are worried about gender-based abuse and systemic harms, and that is effectively an exception for abusers. There are exceptions in it for

32. *Id.* § 2(h)(2)(B).

33. *Id.* §§ 2(h)(4)(A), (B).

34. Barbara Ortutay, *President Trump Signs Take It Down Act, Addressing Nonconsensual Deepfakes. What Is It?*, AP NEWS (May 20, 2025), https://apnews.com/article/take-it-down-deepfake-trump-melania-first-amendment-741a6e525e81e5e3d8843aac20de8615 [https://web.archive.org/web/20260221195607/https://apnews.com/article/take-it-down-deepfake-trump-melania-first-amendment-741a6e525e81e5e3d8843aac20de8615].

35. Take It Down Act, *supra* note 3, § 3(a)(3).

36. *See* 47 U.S.C. § 230(c)(1).

law enforcement and intelligence gathering.[37] But there is also an exception for a person who possesses or publishes a digital forgery of himself or herself, engaged in nudity or sexually explicit conduct.[38] That is to say, if you are a partner with someone at the time that you record the video—and you're in the video with someone else, this provision provides that you are exempt from it. This is precisely the kind of exemption that you might expect to be exploited by abusers.

I'll make one last observation with regards to the dangers associated with empowering federal agencies to enforce law. Consider the Federal Communications Commission's (FCC) recent threats, or the chair of the FCC Brendan Carr's recent threats to ABC and Jimmy Kimmel. Carr argued that the agency's news distortion guidance justified the cancelation of Kimmel's late-night show. When an agency makes threats on the basis of a broadly-worded statute, the risks to protected speech can be great. For the same reason, I think we should worry about the delegations of authority that are too broad. That said, I think there are very few institutions or entities that are capable of addressing the problems I described better than federal agencies. Thank you.

[APPLAUSE]

Responding to a question from David Louk ("My question is for Olivier, and I'm just kind of curious if you're comfortable speculating about what you anticipate will happen in 2026 once the platform obligations go into effect. Because I can see a universe where this starts off a little rocky and then turns out, kind of similar to the Copyright takedown request process, which I think at this point in 2025 is not overly controversial in terms of the broad scope of the way that the process works.

But I could also see either, as you said, the FTC being somewhat opportunistic in the way that it's enforcing it. And I could also see a NetChoice type challenge from one or more platforms to raise the overbreadth issues. Just, to the extent you feel comfortable speculating what you think may happen, I'd be very curious, since we've kind of had this year long period of waiting to find out.")

[SYLVAIN] My speculation is going to be as good as yours. I agree it's subject to manipulation. There is no counter-notification process, as you know. The FTC, the question of whether to go after a platform will be contingent on the FTC's regulatory priorities. And as someone who believes in agencies—believe it or not, I do—this gives me a special concern. And this is not something that is inevitable. Congress could write a law that attends to these problems, but I'm afraid may not have. I don't know what's going to happen, but I think whatever you are guessing, your guess is as good as mine.

37. Take It Down Act, *supra* note 3, §§ 2(h)(3)(C)(i)–(iii).
38. Take It Down Act, *supra* note 3, § 2(h)(3)(C)(iv).

Deepfakes and Private International Law

Edouard Treppoz[*]

Private international law plays a crucial role in addressing deepfakes. By their very nature, deepfakes are inherently international. First, they are digital creations, which means they benefit from ubiquity—the same deepfake can be viewed simultaneously in Paris and New York. Second, deepfakes are primarily disseminated via the internet, making them accessible virtually everywhere. Their global reach necessitates a private international law (PIL) approach. The key question then becomes: What rights are implicated by deepfakes, and how can PIL be applied to each?

As explained in other articles published as part of this symposium, personality rights are clearly central to the issue. Neighboring rights should not be overlooked either, especially considering that AI models may be trained using performances protected under these rights. Finally, from a European Union perspective, regulatory frameworks, such as the AI Act, are also highly relevant. This Article's objective will be to develop a PIL analysis tailored to each of these dimensions. As shown, the methodology differs depending on whether the issue concerns personality rights or regulatory compliance.

I. PERSONALITY RIGHTS AND PIL

Although copyright and even neighboring rights are not central to digital replicas, personality rights appear to play a key role in their analysis. Indeed, these replicas digitally duplicate the voice and the image of a real person, raising the question of authorization. In a domestic case, domestic law usually applies. Because the replica is digital, however, it can be seen and communicated everywhere, meaning that PIL questions must be resolved. Two different sets of questions must be answered. Firstly, which court has jurisdiction? Secondly, which law, or laws, applies?

When it comes to jurisdiction in Europe, an EU regulation applies: Regulation 1215/2012, also called the Brussels I bis Regulation.[1] The goal of this regulation is to

* Professor at the University Paris 1 Panthéon-Sorbonne.

1. *See* Regulation 1215/2012, of the European Parliament and of the Council of 12 December 2012 on Jurisdiction and the Recognition and Enforcement of Judgments in Civil and Commercial Matters, 2012 O.J. (L 351) [hereinafter "Brussels I bis"].

harmonize EU jurisdictional rules. Consequently, the rules determining the competence of a French court are the same as those determining the competence of an Italian court. In both cases, the Brussels I bis Regulation is applicable. Nevertheless, this harmonization is limited to EU cases. If the defendant is not in an EU Member State, the Brussels I bis Regulation is not applicable. As a consequence, the means of determining whether a French court or an Italian court has jurisdiction against a U.S. defendant will depend on French or Italian PIL rules.

This being said, what are the connecting factors under which a French or Italian court may have jurisdiction under the Brussels I bis Regulation, in order to determine whether a digital replica infringes personality rights? The first connecting factor focuses on the domicile of the defendant. If the replica has been created and displayed by a person domiciled in France, French courts have jurisdiction pursuant to article 4 of the Brussels I bis Regulation.[2] Such a scenario could occur, but it is not very common. The second connecting factor is the location of the tort. Pursuant to article 7, paragraph 2 of the Brussels I bis Regulation, "in matters relating to tort, delict or quasi-delict, in the courts for the place where the harmful event occurred or may occur" may have jurisdiction.[3] If the defendant is domiciled in a non-EU country, national PIL rules are applicable. The solution would be the same, since the *forum delicti* is largely recognized. Such a general recognition of the *forum delicti* raises a classical question based on the localization of the tort, especially when committed in a digital manner.

Libel and pollution cases have already shown that the tort at issue could be located in various countries, the place where the harmful act occurred being different from the place where the harm is felt.[4] Moreover, there could be more than one place where the harm is felt. Indeed, the consequences of a defamation could be located in any country where the newspapers are distributed. The European Court of Justice has interpreted article 7, paragraph 2 in a broad manner, explaining that a plaintiff may initiate legal proceedings in either the country where the tortious act occurred or the country where the consequences of the tort are felt , both localizing the tort. The multi-localization of the tort extends the jurisdictional options offered to the plaintiff. Nevertheless, if there is more than one place where the harm is felt, each of the corresponding courts may only have a limited territorial jurisdiction.[5] To put it simply, a French plaintiff may bring a legal matter before an Italian court for defamation in a French newspaper only if the French newspaper is distributed in Italy. But the Italian judge would have jurisdiction only for the harms committed in Italy, not those committed abroad. The damage would thus be limited to the number of newspapers being distributed in Italy. This limitation was at the time justified by a desire to avoid forum shopping,[6] especially in favor of UK courts.

2. *Id.* art. 4.

3. *Id.* art. 7(2).

4. *See, e.g.*, Case C-68/93, Shevill v. Presse Alliance SA, 1995 E.C.R. I-450; Case 21-76, Bier v. Mines de Potasse d'Alsace SA, 1976 E.C.R. 1735.

5. *See Shevill*, 1995 E.C.R. ¶ 33 (ruling that when a libel plaintiff brings suit in a country where the harm occurred, that court has "jurisdiction to rule solely in respect of the harm caused in the State of the court" and not harms that occurred elsewhere).

6. *See id.* ¶ 79.

The question becomes much more complex with the internet. Indeed, a tort committed on the internet may appear to be committed everywhere in the world, at least if accessibility is taken into account. The question becomes how to assign territorial jurisdiction for a tort committed over the internet. In a decisive case, the European Court of Justice went further than for a tort committed in real life.[7] A plaintiff whose personality rights have been infringed online may still benefit from an option between "the courts of the Member State in which the publisher of that content is established" and "the courts of each Member State in the territory of which content placed online is or has been accessible."[8] If the plaintiff chooses a court where the damage is accessible, that court has "jurisdiction only in respect of the damage caused in the territory of the Member State of the court [seized]."[9] This case was the first time the ECJ accepted a localization of the whole damage at the center of interest of the victim. For the Court,

> [A] person who has suffered an infringement of a personality right by means of the internet may bring an action in one forum in respect of all of the damage caused, depending on the place in which the damage caused in the European Union by that infringement occurred. Given that the impact which material placed online is liable to have on an individual's personality rights might best be assessed by the court of the place where the alleged victim has his centre of interests, the attribution of jurisdiction to that court corresponds to the objective of the sound administration of justice.[10]

For the ECJ and for EU law, this is a small revolution. When it comes to jurisdiction, the guiding principle is the forum of the defendant. The plaintiff breaking the social peace must pay the price and seize, in principle, the defendant's jurisdiction. That is why the location of the defendant's domicile is the general rule in the Brussels I bis Regulation and in many national legal systems. The ECJ mentions a "fundamental principle attributing jurisdiction to the courts of the defendant's domicile," with special jurisdiction derogating from that fundamental principle.[11] That's why the ECJ in the *Kronofer* case refused to interpret the "place where the harmful event occurred" as referring to the claimant's domicile.[12] For the Court, the principle being the domicile of the defendant, such an interpretation would be too strong a derogation contrary to the objectives of the Convention. This is no longer true when the harmful event leading to a breach of personality rights occurred on the internet. Indeed, six years after *Kronofer*, the Court reached the exact opposite conclusion in the cases *eDate* and *Martinez*.[13] By localizing the harmful event at the center of interest of the victim, the Court creates a very powerful *forum actoris*.[14] Since the center of interest is usually the

7. Joined Cases C-509/09 & C-161/10, eDate Advert., 2011 E.C.R. I-10302.
8. *Id.* ¶ 52.
9. *Id.*
10. *Id.* ¶ 48.
11. Case C-168/02, Kronhofer, 2004 E.C.R. I-6023, ¶ 13.
12. *Id.* at ¶ 44.
13. Joined Cases C-509/09 & C-161/10, eDate Advert., 2011 E.C.R. I-10302.
14. Jean-Baptiste Racine, *Le* forum actoris *en droit international privé* [*Forum Actoris* in Private International Law], 23 TRAVAUX DU COMITÉ FRANÇAIS DE DROIT INTERNATIONAL PRIVÉ 31 (2019) (Fr.).

habitual residence, the consequence is that the plaintiff may bring the matter to their local court. The home advantage is real. Moreover, the court being seized will have jurisdiction in respect to all the damages caused. Jurisdiction is worldwide. Clearly, the recognition of a *forum actoris* in that case is not neutral from a policy point of view. Political considerations explain such a global localization of the tort at the place of the victim. The internet is seen as a danger, especially when it comes to personality rights. The *forum actoris* is understood as a procedural answer to that danger, a way to compensate the risk created by the internet. This represents a small revolution in Europe, as it shows that jurisdiction rules are fed by policy considerations; but is nothing really new in the United States, since a U.S. court has endorsed the same position in copyright.[15] In the *Penguin* case, the New York Court of Appeals concluded that "[i]n copyright infringement cases involving the uploading of a copyrighted printed literary work onto the Internet . . . the situs of injury for purposes of determining long-arm jurisdiction under [the relevant section of New York's long-arm-jurisdiction statute is] . . . the location of the copyright holder."[16]

However, the following question depends on the applicable law: Once a court's jurisdiction has been established, which law should apply? The question is particularly important when the plaintiff decides to go to court where their center of interest is located. In that case, the court has worldwide jurisdiction, as the whole damage is localized there. Consequently, a French judge may have jurisdiction to determine whether a digital replica is an infringement not only for France but on a worldwide basis. A French court's jurisdiction, indeed, may not necessarily lead to the application of French law. From an EU perspective, two questions must be resolved. First, the question has to be characterized. Second, once the characterization is made, the relevant connecting factor has to be chosen. To complicate matters further, there is no EU harmonization on personality rights' conflict of law rules. The Rome II Regulation which is applicable to IP infringement does not include personality rights.[17] Article 1, paragraph 2, subparagraph (g) says that "non-contractual obligations arising out of violations of privacy and rights relating to personality, including defamation [shall be excluded from the scope of Rome II regulation]."[18] Consequently, a French judge may apply French PIL while an Italian judge may apply Italian PIL. Not only are the PIL systems different in form, but they are also different in substance.

Indeed, conflict of law rules are not the same among EU Member States. Under the Italian Private International Law Act, personality rights may depend on the national law of the person.[19] Pursuant to article 24 of the Act, the existence and content of personality rights are determined by the national law of the person, while the

15. É. FARNOUX, LES CONSIDÉRATIONS SUBSTANTIELLES DANS LE RÈGLEMENT DE LA COMPÉTENCE INTERNATIONALE DES JURIDICTIONS [Substantial Considerations in Regulating International Jurisdiction] (2022).

16. Penguin Grp. (USA) Inc. v. Am. Buddha, 946 N.E.2d 159, 161–62 (N.Y. 2011).

17. *See* Regulation 864/2007, of the European Parliament and of the Council of 11 July 2007 on the Law Applicable to Non-Contractual Obligations, 2007 O.J. (L 199) [hereinafter "Rome II"].

18. *Id.* art. 1(2)(g).

19. Art. 24, Riforma del Sistema Italiano di Diritto Internazionale Privato [Reform of the Italian System of Private International Law], n. 218, May 31, 1995.

consequence of the infringement is determined by the *lex loci deliciti*. Article 24 organizes what PIL specialists call a "dépeçage" between the law of the right and the law of the consequence of the infringement.[20] Such a *dépeçage* could be compared to the *Itar-Tass* case, in which a U.S. judge applied Russian law in order to determine whether the plaintiff possessed rights under Russian law, but applied U.S. law to determine infringement.[21] Let's imagine an unauthorized digital replica of the Senegalese-Italian influencer Khaby Lame. The digital replica has been made by an extremist activist living in Hungary, but who still has a strong connection with Italy. Khaby Lame has his habitual residence in Italy. Let's also assume he is from Senegal and does not have Italian nationality. Based on *Edate*, Lame may bring an action in Italian court, for worldwide damage that is localized in Italy. But the Italian judge would have to apply Senegalese law in order to determine whether Khaby Lame held personality rights. Only Senegalese law would be applicable to determine the content of personality rights.

Under French private international rules, this outcome would be different. Forty years ago, France's highest civil court decided a very important case concerning infringement of personality rights.[22] A French newspaper displayed some unauthorized pictures of Farah Diba, the last wife of the Shah of Iran. The tort took place in France, but the victim was Iranian. The newspaper argued that no infringement had occurred since Iranian law did not recognize personality rights. The Court decided not to apply the plaintiff's national law, ruling that the case required application of the law where the tort had taken place.[23] Thus, under French law, the question becomes how to localize the tort. The fact that, from a jurisdictional point of view, the localization could correspond to the center of interest of the victim may not necessarily lead to the same localization from a conflict of law point of view. Moreover, there is no clear precedent under French law on the question of applicable law for the virtual infringement of personality rights. Nevertheless, some influential scholars have defended a localization at the domicile of the victim.[24] Furthermore, the Project for the French Codification of Private International Law endorses such a localization.[25] Pursuant to article 95, in case of an infringement of personality rights, the law of the center of interest of the injured

20. "Dépeçage" translates literally to "cutting up" or "dismemberment"; in the context of PIL it refers to courts applying the laws of different jurisdictions to separate issues within the same case.

21. Itar-Tass Russian News Agency v. Russian Kurier, Inc., 886 F. Supp. 1120 (S.D.N.Y. 1995).

22. Cour de cassation [Cass.] [supreme court for judicial matters] 1e civ., Apr. 13, 1988, Bull. civ. I, 86-15.524 (Fr.).

23. *Id.*

24. *E.g.*, PIERRE BOUREL, DU RATTACHEMENT DE QUELQUES DÉLITS SPÉCIAUX EN DROIT INTERNATIONAL PRIVÉ [ON THE CONNECTING FACTORS OF CERTAIN SPECIAL TORTS IN PRIVATE INTERNATIONAL LAW] Vol. II, 336–39 (1989); François Dessemontet, *Internet, les droits de la personnalité et le droit international privé* [*Internet, Personality Rights and Private International Law*], 2 MEDIA LEX 77 (1997).

25. PROJET DE CODE DE DROIT INTERNATIONAL PRIVÉ [DRAFT CODE OF PRIVATE INTERNATIONAL LAW] (2022) https://www.justice.gouv.fr/sites/default/files/migrations/textes/art_pix/projet_code_droit_international_prive.pdf [https://web.archive.org/web/20260203214338/https://www.justice.gouv.fr/sites/default/files/migrations/textes/art_pix/projet_code_droit_international_prive.pdf].

person shall apply.[26] The combination of this conflict of law rule and the *Edate* decision creates a very powerful tool in favor of the victim. Let's imagine a digital replica of Catherine Deneuve created in the United States and displayed worldwide. The French actress may bring a legal action for damages in French court. French law would be the only applicable law, even though the replica could be seen everywhere. Such a combined solution is a very powerful tool for the victim.

Finally, the icing on the cake would be the possibility for the presiding judge to order a removal injunction with worldwide effects. Interestingly, the ECJ has said that such an injunction could be ordered by a judge having a worldwide jurisdiction. For the Court, such a "single and indivisible" order may only be issued by a court with jurisdiction on the whole.[27] That is the case for the court of the domicile of the victim. The second condition for a universal order is that the infringement be committed everywhere.[28] As previously explained, a French judge may apply French law if it is the law of the domicile of the victim that localizes the tort committed. Consequently, for a replica of Catherine Deneuve, a French judge may order a removal that would theoretically have extraterritorial effects. Nevertheless, the effectiveness of the tool depends on local means of enforcement. If the French solution could be enforced from France or from the EU, the tool would be very strong. If the decision requires U.S. recognition, the tool might be quite weak. The combination of *eDate* and article 95 of the French Project for PIL Codification constitutes a theoretically strong tool whose practical effectiveness may vary across cases.

II. TERRITORIAL SCOPE OF APPLICATION OF THE AI ACT

If personality rights are clearly at the center of the digital replica issue, we have also seen from an EU point of view that regulation is also highly important. With the GDPR, the DSA, the DMA, and more recently the AI Act, the EU Commission has drafted numerous texts in order to regulate digital activity.[29] Two common features can

26. *Id.* at art. 95.

27. C-194/16, Bolagsupplysningen and Ilsjan, ECLI:EU:C:2017:766, ¶ 48 (Oct. 17, 2017) ("[I]n the light of the ubiquitous nature of the information and content placed online on a website and the fact that the scope of their distribution is, in principle, universal, an application for the rectification of the former and the removal of the latter is a single and indivisible application and can, consequently, only be made before a court with jurisdiction to rule on the entirety of an application for compensation for damage pursuant to the case-law resulting from the judgments of 7 March 1995, *Shevill and Others* (C-68/93, EU:C:1995:61, paragraphs 25, 26 and 32), and of 25 October 2011, *eDate Advertising and Others* (C-509/09 and C-161/10, EU:C:2011:685, paragraphs 42 and 48), and not before a court that does not have jurisdiction to do so.") (citation omitted).

28. *Cf.* C-235/09, DHL Express France v. Chronopost, 2011 E.C.R. I-2825, ¶ 45, *and* C-18/18, Glawisching-Piesczek, ECLI:EU:C:2019:821, ¶ 50 (Oct. 3, 2019) ("Consequently, and also with reference to paragraphs 29 and 30 above, Directive 2000/31 does not preclude those injunction measures from producing effects worldwide."); Regulation 2022/2065, of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC [hereinafter "Digital Services Act"], recital 31 ("Therefore, this Regulation does not provide the legal basis for the issuing of such orders, nor does it regulate their territorial scope or cross-border enforcement.").

29. See Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 [hereinafter "GDPR"]; Digital Services Act (DSA), *supra* note 28; Regulation 2022/1925, of the European

be identified. First, the goal of these texts is not to sanction *ex post*, but to impose a behavior *ex ante*. Compliance is key for the EU Commission in order to protect the EU digital market. Second, all these texts share the same means of determining their territorial scope of application. Those rules are unilateral, meaning that they only determine situations in which EU regulations apply, not those in which non-EU laws apply.[30] This being said, these unilateral conflict of law rules endorse two cumulative connecting factors in order to determine their territorial scope of application.

Pursuant to article 3 of the GDPR, the Regulation "applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not."[31] An EU controller must comply with the GDPR wherever the processing takes place. The goal is to avoid any circumvention of EU law by manipulating the place of processing. The specificity of the GDPR and other EU regulations on digital activity territorial scope serves to add another connecting factor when the processor or controller does not have any establishment in the EU. In such a situation, the GDPR may still be applicable if data subjects are in the Union, goods or services are offered in the Union, or if there is a monitoring of behavior taking place within the Union. A U.S. startup with no physical presence in the EU must nevertheless comply with the GDPR if the startup offers goods in the EU.

This dual and subsidiary scope of application of the GDPR becomes a common trend for all EU regulations concerning digital activities. Although the relevant provisions of the AI Act are not drafted in identical terms, the result appears to be largely the same. First, deployers of AI systems that are established or located within the EU must comply

---

Parliament and of the Council of 14 September 2022 on Contestable and Fair Markets in the Digital Sector and amending Directives 2019/1937 and 2020/1828 (Digital Markets Act, or DMA), 2022 O.J. (L 265); Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence and amending Regulations 300/2008, 167/2013, 168/2013, 2018/858, 2018/1139 and 2019/2144 and Directives 2014/90, 2016/797 and 2020/1828 [hereinafter "EU AI Act"], 2024 O.J. (L 1689) .

30. S. Francq, *Universalisme Versus Bilatéralisme: Une Opposition Ontologique ou un Débat Dépassé? Quelques Considérations de Droit Européen sur un Couple en Crise Perpétuelle* [Universalism Versus Bilateralism: An Ontological Opposition or an Outdated Debate? European Law Considerations on a Couple in Perpetual Crisis], *in* QUEL AVENIR POUR LA THÉORIE GÉNÉRALE DES CONFLITS DE LOIS? 51 (T. Azzi & O. Boskovic eds., 2015).

31. GDPR, *supra* note 29, art. 3(1); *see, e.g.*, M. E. Ancel, *D'une Diversité à L'autre: A propos de la "Marge de manœuvre" Laissée par le Règlement Général sur la Protection des Données aux États Membres de l'Union Européenne* [*From One Diversity to Another: About the "Margin of Maneuver" Left by the General Data Protection Regulation to the Member States of the European Union*], REVUE CRITIQUE DE DROIT INTERNATIONAL PRIVÉ [REV. CRI. DR. INT. PRIV.] 647 (2019) (Fr.); Ludovic Paillier, *L'Applicabilité Spatiale du Règlement Général sur la Protection des Données: Commentaires de l'Article 3* [*The General Data Protection Regulation's Territorial Applicability: Commentaries on Article 3*], JOURNAL DU DROIT INTERNATIONAL (CLUNET) 823 (2018) (Fr.); C. Kohler, *Conflicts of Laws Issues in the 2016 Data Protection Regulation of the European Union*, 52 RIVISTA DI DIRITTO INTERNAZIONALE PRIVATO E PROCESSUALE [RIV. DIR. INT. PRIV. PROC.] 653 (2016) (It.); Luís de Lima Pinheiro, *Law Applicable to Personal Data Protection on the Internet: Some Private International Law Issues*, ANUARIO ESPAÑOL DE DERECHO INTERNACIONAL PRIVADO [AN. ESP. DER. INT. PRIV.] 163, 163 (2018); PEDRO DE MIGUEL ASENSIO, CONFLICT OF LAWS AND THE INTERNET 113–96 (2020).

with the AI Act.[32] Under this rule, for example, the French unicorn Mistral must comply with the AI Act. Much as under the GDPR, a provider established outside of the EU must respect the AI Act if he introduces AI systems in the EU market. The same is true where the output produced by the AI system is used within the Union. An Indian provider must thus comply with the AI Act if it targets the EU market by placing an AI system in the EU or just by offering output that is used there.

The underlying objective is the same: to ensure that the policy aims pursued by these Regulations are achieved for the benefit of EU citizens and of the EU market. This dual, subsidiary scope of application, based on the place of establishment and on the target identified, seeks to prevent circumvention of these laws via manipulation of either. In the end, what counts is whether the EU public is targeted. Where that is the case, EU law must be respected. Criticism has nevertheless been directed at this EU approach. Part of that criticism rests on the allegedly extraterritorial scope of these Regulations.[33] That objection is not particularly persuasive, as extraterritoriality is an inherent principle in PIL. PIL indeed applies when a legal issue can be localized in more than one country. The conflict of law rule designates a single law, meaning that this one legal system will be applicable to a situation not solely confined to the territory of a single State. From this perspective, extraterritoriality is not a problem for PIL, but part of the solution.[34] A more substantial concern, instead, lies in the potentially exorbitant scope of application of these Regulations.[35] When the territorial scope of a norm is determined by one single connecting factor, the factor may be susceptible to "bilateralization," that is, capable of giving rise to bilateral, reciprocal relationships between two states.

As an example, the French copyright rules on satellite broadcasting are applicable provided that the uplink is located in France.[36] This rule could be "bilateralizable," meaning that if the uplink were in Germany, German law would be applicable. A unilateral conflict rule may be "bilateralizable" when only one connecting factor is used. Since only one connecting factor is used, domestic law and foreign law are on equal footing. Consequently, the scope of EU regulations could be considered extensive, because there is more than one connecting factor. Moreover, the absence of any bilateral conflict of law rule leads to a risk under which dual and contradictory order may exist.[37] An Indian AI provider located in India may have to respect Indian law if

32. EU AI Act, *supra* note 29, art. 2.

33. *E.g.*, J. Heymann, *L'extraterritorialité du Droit International Privé Européen* [*Extraterritoriality in European Private International Law*], *in* L'EXTRATERRITORIALITÉ EN DROIT DE L'UNION EUROPÉENNE 69, 86 (E. Dubout, F. Martucci & F. Picod eds., 2021); Céline Castets-Renard, *Extraterritorialité du Droit Européen des Activités Numériques* [*The Extraterritoriality of European Digital Law*], *in* L'EXTRATERRITORIALITÉ EN DROIT DE L'UNION EUROPÉENNE 113, 119 (E. Dubout, F. Martucci & F. Picod eds., 2021).

34. M. E. Ancel, *supra* note 31 at 647; Sabine Corneloup, *Commentary on the Book "Solving the Internet Jurisdiction Puzzle,"* REV. CRI. DR. INT. PRIV. 404 (2018).

35. M. Audit, *Les Lois Extraterritoriales Américaines Comme Facteur D'accélération de la Compliance* [*U.S Extraterritorial Laws as An Accelerating Factor for Compliance*], *in* COMPLIANCE: L'ENTREPRISE, LE RÉGULATEUR ET LE JUGE 45, 47 (N. Borga, J. C. Marin & J. C. Roda eds., 2018).

36. Code de la Propriété Intellectuelle (C.P.I.) 4, art. L. 122-2-4 [hereinafter "French Intellectual Property Code"].

37. Louis d'Avout, *L'entreprise et les Conflits Internationaux de Lois* [Business and International Conflicts of Laws], RCADI, 811, 811–13 (2019).

the application of Indian law is based on its establishment. If this Indian provider puts AI systems on the EU market, the AI Act would also be applicable. Because both laws would apply to the same activity, the Indian provider may have to respect contradictory orders from India and the EU.

Going back to the AI Act, article 50, paragraph 4 requires that any deployer of AI systems generating deepfakes disclose that the content has been artificially generated or manipulated.[38] This obligation applies if the deployer is in the EU, if it places AI systems in the EU, or if the output is used in the EU.[39] The same is true for article 53, paragraph 1(c).[40] Under this provision, providers must put in place a policy to comply with Union law on copyright and related rights. It has been explained that a deepfake as such does not constitute an infringement of copyright law nor a related right. Even if the output itself is not an infringement, it is nevertheless clear that an AI system creating deepfakes of Catherine Deneuve would have been trained on interpretations of Catherine Deneuve. Related rights that are inapplicable to the output become applicable to the training process. In order to comply with the AI Act, then, the provider must comply with EU policy. More simply, authorization could be required for interpretations for which the opt-out has been correctly exercised. Based on article 2, the AI Act is applicable if the output is used in the EU. Therefore, Indian providers have to respect EU copyright law if the deepfake produced by the AI system is used in the EU. Recital 106 is even clearer, explaining that "any provider placing a general purpose AI model on the Union market should comply with this obligation, regardless of the jurisdiction in which the copyright-relevant acts underpinning the training of those general-purpose AI model take place."[41] The Indian providers must comply with EU copyright law even if the training is done in Japan, where it is copyright free.[42] Under Japanese copyright law, training does not require any authorization. Nevertheless, because the Indian provider places an AI system on the Union market, he has to comply with EU law and training is no longer free. If he has not put in place a policy in order to comply with EU copyright law, it will be considered as a breach of the AI Act since article 52's requirement for territorial applicability is not respected.

What is unclear is whether the provider's actions could also be characterized as copyright infringement. The scope of the application of the AI Act is unilateral and

38. EU AI Act, *supra* note 29, art. 50(4).

39. *Id.* recital 22 (pursuant to recital 22, the term used should be understood as being "intended to be used . . . [t]o prevent the circumvention of this Regulation and to ensure an effective protection of natural persons located in the Union, this Regulation should also apply to providers and deployers of AI systems that are established in a third country, to the extent the output produced by those systems is intended to be used in the Union.").

40. *Id.* art. 53(1)(c).

41. *Id.* recital 106.

42. Chosakuken Hō [Copyright Law], Law No. 48 of 1970, art. 30-4 (Japan), as amended by Law No. 30 of 2018, English translation available at https://www.cric.or.jp/english/clj/cl2.html [https://web.archive.org/web/20260210141600/https://www.cric.or.jp/english/clj/cl2.html] (pursuant to article 30-4 of the Japanese law, no authorization is required for TDM); Tatsuhiro Ueno, *The Flexible Copyright Exception for "Non-Enjoyment" Purposes—Recent Amendment in Japan and Its Implication*, 70(2) GRUR INT'L 142 (2021).

extremely broad. It does not align with the scope of application of copyright law, which is determined by a bilateral conflict-of-law rule based on the country in which the protection is sought.[43] Consequently, there is no symmetry between the scope of application of the AI Act and the scope of application of copyright law. This means that not complying with EU copyright rules could be a breach of the AI Act, but not necessarily an infringement under copyright law if the training occurred in Japan.

The question becomes whether EU copyright law applies to a deepfake generated in the EU but by an AI tool trained in Japan. The applicable law would be that of the country in which protection is sought. It is unclear, however, whether localization has to be made in a distributive manner, act by act, or whether it should be made in a more holistic manner, taking into account not each act but the infringing activity as such.[44] If one considers that the object to be localized is the infringing activity as a whole, the activity occurs in Japan and in the EU.[45] It is then clear that the law of the public being targeted has to be applicable, based on strong EU case law.[46] Following this analysis, EU law would be applicable to the whole activity. Consequently, the provision of

43. *See* Cour de cassation [Cass.] [supreme court for judicial matters] 1e civ., Apr. 10, 2013, Bull. civ. I, 11-12.508 (Fr.) (characterizing the article 5 par 2 of the Berne Convention as a bilateral conflict of law rule) (contra 2 SAM RICKESTON & JANE C. GINSBURG, INTERNATIONAL COPYRIGHT AND NEIGHBOURING RIGHTS: THE BERNE CONVENTION AND BEYOND, 1299 (2006)).

44. Pierre Bourel, LES CONFLITS DE LOIS EN MATIÈRE D'OBLIGATIONS EXTRACONTRACTUELLES [CONFLICT OF LAWS IN TERMS OF EXTRACONTACTUAL OBLIGATIONS] 65–66 (1961) (distinguishing between one harmful event dissociated and a parcel of harmful events and arguing in favor of an economic localization of the harmful event); André Lucas, ASPECTS DE DROIT INTERNATIONAL PRIVÉ DE LA PROTECTION D'ŒUVRES ET D'OBJETS DE DROITS CONNEXES TRANSMIS PAR RÉSEAUX NUMÉRIQUES MONDIAUX [PRIVATE INTERNATIONAL LEGAL ASPECTS OF THE PROTECTION OF WORKS AND OBJECTS OF RELATED RIGHTS TRANSMITTED THROUGH DIGITAL NETWORKS], 26 (1998), https://www.wipo.int/edocs/mdocs/mdocs/fr/gcpic/gcpic_1.pdf [https://web.archive.org/web/20260221215814/https://www.wipo.int/edocs/mdocs/mdocs/fr/gcpic/gcpic_1.pdf].

45. For a case clearly in favor of this approach, *see* Joined Cases C-24/16 & C-25/16, Nintendo Co. Ltd. v. BigBen Interactive GmbH, ECLI:EU:C:2017:724, ¶ 109 (Sept. 27, 2017) ("As regards the second set of circumstances mentioned by the referring court, by which it inquires about the law applicable when an operator has goods that allegedly infringe the rights protected by a Community design shipped by a third-party undertaking to a Member State other than the one in which it is domiciled, it should be noted, as stated in paragraph 103 of the present judgment, that the correct approach for identifying the event giving rise to the damage within the meaning of Article 8(2) of Regulation No 864/2007 *is not to refer to each of a defendant's alleged acts of infringement, but to make an overall assessment of that defendant's conduct* in order to determine the place where the initial act of infringement at the origin of that conduct was committed or threatened by it.") (emphasis added). On that question, *see* Edouard Treppoz, *Le Paradoxe du Principe de Territorialité en Droit Européen de la Propriété Intellectuelle, in* LE DROIT À L'ÉPREUVE DES SIÈCLES ET DES FRONTIÈRES, MÉLANGES EN L'HONNEUR DE B. ANCEL 1510 (2018).

46. Case C-324/09, L'Oréal, SA v. eBay Int'l A.G., ECLI:EU:C:2011:474 (July 12, 2011) (seminal case); *see also* Case C-76/24, Tradeinn Retail Servs. S.L. v. PH, ECLI:EU:C:2025:593, ¶ 37 (Aug. 1, 2025) (explaining that "[a]s is apparent from paragraph 34 above, the proprietor of a trade mark may prohibit a third party from offering, inter alia by means of online advertising, goods under that sign notwithstanding the fact that that third party, the server of the website which it uses or those goods are located outside the Member State of registration, if that offer is targeted at consumers in the territory of that Member State. In such a situation, that proprietor is also entitled to prohibit that third party from stocking those goods outside that territory if that stocking constitutes a preliminary step to the making of such an offer or its implementation, with the result that it may be regarded as having been carried out for that purpose."). For more details, *see* Treppoz, *supra* note 45.

deepfakes in the EU, generated by an AI system trained in Japan, would constitute an infringement of EU law.[47] If localization must be determined on an act-by-act basis, it becomes necessary to identify whether the reproduction right or the right of communication to the public is applicable.[48] A recent trend in the EU considers that the right of communication to the public is applicable.[49] In that case, the reasoning is the same: the act is localized in Japan for as to the place where the harmful act occurred and in the EU as to the place where the harm is felt. The same case law applies, and EU law governs as the law of the public who is being targeted. If, by contrast, only the reproduction right was implicated, and localization likewise had to be determined act by act, Japanese law would apply. A dissymmetry would then arise: the AI provider would fall outside of the AI Act, yet he would not be liable as an infringer. Nevertheless, it may be possible to argue, under the Japanese law otherwise applicable, that EU law should apply as an overriding mandatory rule, since its territorial scope of application is determined by recital 105.[50] Under EU law, "overriding mandatory provisions are provisions the respect for which is regarded as crucial by a country for safeguarding its public interests, such as its political, social or economic organization, to such an extent that they are applicable to any situation falling within their scope, irrespective of the law otherwise applicable."[51] Recital 105 clearly expresses the public interest at stake and the willingness to safeguard a political and economic organization in the EU by requiring the application of EU copyright law if outputs are generated in the EU (irrespective of the place of training).[52] the use of overriding mandatory rules in copyright law is not unprecedented: it has already been accepted in relation to moral

47. Eleonora Rosati, *Infringing AI: Liability for AI-Generated Outputs under International, EU, and UK Copyright Law*, 16 EUR. J. RISK REGUL. 603 (2025).

48. Alexander Peukert, *Regulating IP Exclusion/Inclusion on a Global Scale: The Example of Copyright vs. AI Training* ¶ 11 (Goethe Univ. Frankfurt Fac. of L., Research Paper No. 3, 2024) ("If AI training were subject only to conventional copyright laws, AI providers could still operate effectively, despite the uncertainties just described. The reason is that according to the universally accepted principle of territoriality and the corresponding rule of *lex loci protectionis*, reproductions in the course of AI model training are governed solely by the copyright law of the country in which these reproductions take place"); following the same position, *see* JOÃO PEDRO QUINTAIS, COPYRIGHT, THE AI ACT AND EXTRATERRITORIALITY 13 (2025), http://dx.doi.org/10.2139/ssrn.5316132 [https://perma.cc/EC48-A5TC].

49. *See* EXPERT GROUP ON COPYRIGHT & ARTIFICIAL INTELLIGENCE, RECOMMENDATIONS FROM EXPERT GROUP ON COPYRIGHT & ARTIFICIAL INTELLIGENCE 65 (Danish Ministry of Culture, Sep. 2025), https://www.aepo-artis.org/wp-content/uploads/2025/10/1.-Danish-Copyright-expert-group-report-EN.pdf, [https://web.archive.org/web/20260311171401/https://www.aepo-artis.org/wp-content/uploads/2025/10/1.-Danish-Copyright-expert-group-report-EN.pdf].

50. EU AI Act, *supra* note 29, recital 105.

51. Council Regulation 593/2008, art. 9(1), 2008 O.J. (L 177) 6, 13 (on the law applicable to contractual obligations (Rome I)).

52. EU AI Act, *supra* note 29, recital 105.

rights[53] and remuneration rights.[54] This supports the argument that EU law may apply to training carried out in Japan, even where the deepfake is generated in the EU. More fundamentally, it shows how PIL is no longer a neutral technique, but rather a framework made of deliberate political choices.

(October 2025)

53. Cour de cassation [Cass.] [supreme court for judicial matters] 1e civ., May 28, 1991, Bull civ. I, 89-19.522 (Fr.) (applying French law to the ownership of moral rights for an American movie, whereas the applicable conflict of law rule designated U.S. law as being the law of the origin of the work. The direct application of French law is justified since moral rights are characterized as being "lois d'application impérative."). On that decision, *see* Jane C. Ginsburg & Pierre Sirinelli, *Authors and Exploitations in International Private Law: The French Supreme Court and the Huston Film Colorization Controversy*, 15 COLUM. J.L. & ARTS 135 (1991).

54. Article L. 131-24 of the French Intellectual Property Code, creating a super "loi de police" trumping the choice of law clause and the choice of court agreement (*see* Edouard Treppoz, *Le Droit Contractuel des Auteurs—Aspects de Droit International Privé* [*Contractual Authors' Right—Aspects of International Private Law*], 80 PROPRIÉTÉS INTELLECTUELLES 60 (2021) (Fr.). For more general information on authors' remuneration and applicable law, *see* Jane C. Ginsburg & Pierre Sirinelli, *Private International Law Aspects of Author's Contracts: the Dutch and French Examples*, 39 COLUM. J.L. & ARTS 171 (2015).

Deepfakes and Private International Law

Graeme B. Dinwoodie*

* Global Professor of Intellectual Property Law and University Distinguished Professor, IIT Chicago-Kent College of Law; Visiting Professor of Law, Northwestern University Pritzker School of Law.

As this Symposium has shown, the problems for intellectual property law caused by the creation and distribution of so-called "deepfakes" are many and the solutions to those problems remain speculative or, at best, inchoate. But any of the most plausible responses, all of which are developing at the national (or regional, as in the European Union) level, will inevitably raise difficult issues of private international law. In this Article, in assessing the likely role of private international law in resolving legal claims prompted by deepfakes, I stress the broader context in which those questions ought to be considered, building in part on several conceptual propositions contained in Edouard Treppoz's article on how such matters would be handled under EU law.[1] Then, I assess how U.S. courts might approach similar issues of private international law, arguing that existing U.S. rules are inadequate in a number of respects, hampered both by truncated treatment of choice of law in past cases involving publicity rights and outdated (but recently re-discovered) commitments by the U.S. Supreme Court to conduct-based notions of territoriality.

I. THE ROLE OF PRIVATE INTERNATIONAL LAW

A. THE BACKGROUND INTERNATIONAL PICTURE

Professor Treppoz starts from the observation that deepfakes are "inherently international."[2] As a result, he says that private international law plays a crucial role in addressing the problem of deepfakes. This statement clearly echoes claims made thirty years ago with the emerging social and commercial exploitation of the internet. That development prompted the first sustained attention by scholars, courts, and policymakers to the private international law of intellectual property,[3] as well as more radical forecasts about the development of governing rules detached from territorial

1. Eduoard Treppoz, *Deepfakes and Private International Law*, 49 COLUM. J.L. & ARTS 791 (2026)
2. *Id.* at 791.
3. *See* P.B. Carter, *Decisions of British Courts During 1990*, 61 BRIT. Y.B. INT'L L. 386, 401–02 (1991) (noting lack of private international law of intellectual property).

jurisdictions.[4] In fact, the importance of private international law for the deepfakes problem will be far greater, and the challenge far more difficult, than we faced thirty years ago with then-imminent digital copyright and trademark disputes. This is because of the different legal background that existed internationally in 1995.

Private international law is only one of various mechanisms by which territorial rights are reconciled. The system of *public* international intellectual property law contains a range of mechanisms that constrain—rather than referee between—variation in national norms. The extent to which those mechanisms exist will render the role of private international law relatively less—or more—influential.

With the emergence of digital copyright in the 1990s, we saw prompt recourse to the public international copyright system.[5] The resulting treaties (the World Intellectual Property Organization (WIPO) Copyright Treaty and the WIPO Performances and Phonograms Treaty) were neither comprehensive nor detailed.[6] But countries did come together on basic principles.[7] Those treaty commitments were soon bolstered with internationalization through private ordering, by takedown procedures implemented by platforms that transcended national borders.[8] That cross-border private ordering was constructed with the scaffolding of national legislative interventions offering conditional immunity to internet intermediaries (the Digital Millennium Copyright Act and the E-Commerce Directive).[9]

Despite the challenge to territorial rights presented by online uses of trademarks, claims that arose implicated long-standing substantive principles on which there was extensive international agreement. And any uncertainty created by the cross-border character of ubiquitous online uses was ameliorated by common principles of restraint contained in innovative soft law initiatives at WIPO, which quickly took root in parallel national jurisdictions.[10] Addressing the particular scourge of cybersquatting

4. *Cf.* David R. Johnson & David Post, *Law and Borders: The Rise of Law in Cyberspace*, 48 STAN. L. REV. 1367, 1367 (1996) (arguing that a body of laws separate from that which applied offline would need to arise to regulate cyberspace in part through new institutions not tied to traditional national lawmakers).

5. *See* Graeme Dinwoodie, *The WIPO Copyright Treaty: A Transition to the Future of International Copyright Lawmaking?*, 57 CASE W. L. REV. 751 (2007).

6. WIPO Copyright Treaty, Dec. 20, 1996, 2186 U.N.T.S. 121; WIPO Performances and Phonograms Treaty, Dec. 23, 1996, 2186 U.N.T.S. 203.

7. *See* Pamela Samuelson, *The U.S. Digital Agenda at WIPO*, 37 VA. J. INT'L L. 369, 371–72 (1997); Jane C. Ginsburg, *International Copyright: From a "Bundle" of National Copyright Laws to a Supranational Code?*, 47 J. COPYRIGHT SOC'Y 265, 270–72 (2000).

8. Graeme B. Dinwoodie, *Private Ordering and the Creation of International Copyright Norms: The Role of Public Structuring*, 160 J. INSTIT. & THEORETICAL ECON. 161, 169–74 (2004).

9. *See* Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860 (1998) (codified at 17 U.S.C. §§ 512); Council Directive 2000/31/EC on Certain Legal Aspects of Information Society Services, in Particular, Electronic Commerce, in the Internal Market, arts. 12–15, 2000 O.J. (L 178) 1, 12–13.

10. *See* Joint Recommendation Concerning Provisions on the Protection of Marks, and Other Industrial Property Rights in Signs, on the Internet, adopted by Assembly of the Paris Union for the Protection of Industrial Property and General Assembly of the World Intellectual Property Organization, at 2, WIPO Doc. 845(E) (Oct. 2001) (principles for determining whether "use of a sign on the Internet has contributed to the acquisition, maintenance or infringement of a mark or other industrial property right in the sign, or whether such use constitutes an act of unfair competition.").

was one instance where prognostications about rules detached from territorial jurisdictions proved accurate.[11]

That is to say, the treatment of inherently cross-border problems is commonly a mix of private international law, public international law, and forms of private ordering that sideline the formal principle of territoriality.[12] In the 1990s, it was not always clear that these other institutions would step up and solve the problem. As a result, we saw an intensified effort to think through rules of private international law applicable to cross-border intellectual property disputes: first in the failed Hague Convention, then the American Law Institute ("ALI") Principles project in the United States and the counterpart Conflict of Laws in Intellectual Property ("CLIP") project in Europe.[13] The number of fully adjudicated private international law disputes in copyright and trademark law since that activity has been surprisingly small. But this is because the problems have been ameliorated by the early adoption of similar substantive copyright and trademark norms, jumpstarted by the public international process, and extensive cross-border private ordering.

The picture with the deepfakes dilemma is quite different in ways that make the challenge for private international law far more substantial and far more important than it was thirty years ago. This is because even if there is some international agreement on the basic contours of the problem, we do not even share conceptual starting points for the intellectual property part of the solution. This lack of a shared conceptual lens through which to view matters makes agreement on substantive international norms unlikely. Unlike copyright or trademark law in 1996, there are no basic international norms on personality or image protection to extend and adapt to the new factual setting.

---

11. *See generally* Laurence R. Helfer & Graeme B. Dinwoodie, *Designing Non-National Systems: The Case of the Uniform Domain Name Dispute Resolution Policy*, 43 WM. & MARY L. REV. 141 (2001) (outlining the non-national character of the Uniform Domain Name Dispute Resolution Policy).

12. *See generally* Graeme B. Dinwoodie, *The Architecture of the International Intellectual Property System*, 77 CHI.-KENT L. REV. 993, 1011 (2002) (describing the different mechanisms that intellectual property law had developed for addressing cross-border problems).

13. INTELLECTUAL PROPERTY: PRINCIPLES GOVERNING JURISDICTION, CHOICE OF LAW, AND JUDGMENTS IN TRANSNATIONAL DISPUTES (A.L.I. 2008) [hereinafter "ALI PRINCIPLES"]; *See also* HAGUE CONFERENCE ON PRIVATE INT'L LAW, PRELIMINARY DRAFT CONVENTION ON INTERNATIONAL JURISDICTION AND FOREIGN JUDGMENTS IN CIVIL AND COMMERCIAL MATTERS (Oct. 30, 1999), http://www.hcch.net/upload/wop/jdgm2001draft_e.pdf [https://web.archive.org/web/20240623132635/https://assets.hcch.net/upload/wop/jdgm2001draft_e.pdf] ; EUROPEAN MAX PLANCK GRP. ON CONFLICT OF L. IN INTELL. PROP., CONFLICT OF LAWS IN INTELLECTUAL PROPERTY: THE CLIP PRINCIPLES AND COMMENTARY (Oxford Univ. Press 2013) [hereinafter "CLIP PRINCIPLES"]; *see generally* Rochelle Cooper Dreyfuss, *An Alert to the Intellectual Property Bar: The Hague Judgments Convention*, 2001 U. ILL. L. REV. 42 (2001); Graeme B. Dinwoodie, *Developing a Private International Intellectual Property Law: The Demise of Territoriality?*, 51 WM. & MARY L. REV. 711, 719 (2009) ("Those efforts [to negotiate a Hague Convention on Jurisdiction and Recognition of Judgments] floundered in 2000–2001, in large part due to disagreement over how to handle intellectual property cases, forcing the Conference to scale back its efforts and concentrate on a convention validating exclusive choice of court clauses in business-to-business contracts."); Rochelle C. Dreyfuss & Jane C. Ginsburg, *Draft Convention on Jurisdiction and Recognition of Judgments in Intellectual Property Matters*, 77 CHI.-KENT L. REV. 1065 (2002).

Of course, divergence among substantive national rules is why we have private international law; it allows genuine differences to persist and conflicts to be resolved through (supposedly) value-neutral procedural rules. But the lack of a common conceptual understanding here is quite deep. The odd assortment of theories through which intellectual property rights against deepfakes might be asserted under different national laws *also* creates difficulties for private international law. Any sophisticated interest analysis of any choice of law issue in this field is likely to generate far more true conflicts than arose in the realms of copyright or trademark. As just one illustration, it is clear that publicity rights (sometimes called image rights or personality rights) will be at the forefront of deepfake-related intellectual property claims. But the leading private international law projects disagree on whether they even speak to such personality-grounded rights: The ALI Principles purport to apply to right of publicity, while the CLIP Principles explicitly exclude such claims from their scope.[14] Private international law applicable to deepfake claims is operating in a much different space than copyright and trademark laws were in the 1990s.

B. THE DUALITY OF (LIKELY) RELEVANT LEGISLATION

One of the other challenges for private international law in assessing deepfake-related claims is that, based on the current legislative outlook, it will have to tackle two different forms of relevant law. First, we have public law regulation, typified by the AI Act in Europe (but with likely counterparts in the United States).[15] Second, we have

14. *See* ALI PRINCIPLES, *supra* note 13, § 102, cmt. b ("it may even become appropriate for courts to apply the Principles in a case in which not all countries in the world recognize the right claimed" and giving as a leading example the "right of publicity" and which in many European countries under the name "right to one's own image); CLIP PRINCIPLES, *supra* note 13, art. 1:101.C03 ("[T]he right to a person's voice or image, sometimes designated as the right of publicity, is not included in the scope of Article 1:101. While such rights could also be classified as property rights and are in fact often exploited in a similar commercial way, the group did not want to interfere with the prevailing approach in Europe, which considers the protection of these rights and of personality rights in general as a matter of the law of tort or delict.").

15. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 2024 O.J. (L 1689) [hereinafter "EU AI Act"]. The U.S. approach remains in flux. *See* WHITE HOUSE, A NATIONAL POLICY FRAMEWORK FOR ARTIFICIAL INTELLIGENCE (Mar. 19, 2026), https://www.whitehouse.gov/wp-content/uploads/2026/03/03.20.26-National-Policy-Framework-for-Artificial-Intelligence-Legislative-Recommendations.pdf [https://web.archive.org/web/20260325220307/https://www.whitehouse.gov/wp-content/uploads/2026/03/03.20.26-National-Policy-Framework-for-Artificial-Intelligence-Legislative-Recommendations.pdf]. This is also true elsewhere. *See* H.M. GOVERNMENT, REPORT ON COPYRIGHT AND ARTIFICIAL INTELLIGENCE (U.K.) (Mar. 18, 2026), https://assets.publishing.service.gov.uk/media/69ba692226909a14239612e4/CP2602959_-_Report_on_Copyright_and_Artificial_Intelligence_web.pdf [https://web.archive.org/web/20260325182539/https://assets.publishing.service.gov.uk/media/69ba692226909a14239612e4/CP2602959_-_Report_on_Copyright_and_Artificial_Intelligence_web.pdf].

private causes of action that look, with varying degrees of resemblance, more like claims based on traditional intellectual property rights.

This makes private international law analysis more complex. The geographic scope of any public law regulation will be determined differently than the way courts assess the law applicable to a private IP infringement claim. At the very least, U.S. courts are unlikely to directly enforce norms contained in any foreign public law regulation (such as the EU AI Act), though they may in certain circumstances assume jurisdiction over foreign IP law claims.[16] With that said, if U.S. courts engage in a full-blown modern interest analysis to determine applicable law, existence of parallel regulatory regimes may well inform how they define the relevant governmental interests.[17]

C. TERRITORIALITY VERSUS EXTRATERRITORIALITY

It is clear that many of these new laws will (quite intentionally) have some extraterritorial effect. For example, the EU AI Act will impose obligations on those providers located in the EU as well as those outside the EU who target the EU.[18] But I strongly endorse Professor Treppoz's observation that critiques of the EU AI Act based on the allegedly extraterritorial scope of the Regulation are not convincing since, as he puts it, "extraterritoriality is not a problem for [private international law] but part of the solution."[19]

This is a bold assertion, especially when working close to the field of intellectual property where territoriality has an almost talismanic quality.[20] But it is surely correct. Extraterritoriality is inevitable in this context.[21] It is unhelpful to reflexively apply the label of "extraterritoriality" to disapply national laws with unavoidable intrusions on competing sovereignties. We need to view these questions in the context of transborder activity that might render a country's sovereign choices ineffective by making it impossible to effectively prescribe within one's own borders.

---

16. *See* Philip J. McConnaughay, *Reviving the "Public Law Taboo" in International Conflict of Laws*, 35 STAN. J. INT'L L. 255 (1999).

17. Bernhard v. Harrah's Club, 546 P.2d 719, 720–725 (Cal. 1976).

18. *See* EU AI Act, *supra* note 15, art. 2(1)(c) (applying to providers and deployers of AI systems located in a third country where the output produced by the system is used in the EU).

19. *See* Treppoz, *supra* note 1, at 798.

20. *See* Dinwoodie, *supra* note 13 at 714–15 ("Territoriality is a principle that has always received excessive doctrinal purchase in intellectual property law. One can adhere to the basic premises that underlie territoriality without supporting the full range of rules of intellectual property law that are said to reflect the principle."); Rochelle Dreyfuss, *The ALI Principles on Transnational Intellectual Property Disputes: Why Invite Conflicts?*, 30 BROOK. J. INT'L L. 819, 848 (2005) ("In an early presentation of the ALI Principles to the Advisers, a prominent jurist argued that there was no need for choice of law rules because the territorial principle was so obviously applicable.").

21. Dinwoodie, *supra* note 13 at 715 ("Contemporary multi-territorial intellectual property disputes are characterized by an excess of shared but weaker prescriptive and adjudicatory authority. The Article suggests a restrained concept of territoriality that reflects that reality, drawing in particular from the treatment of extraterritoriality in trademark law.").

That is, there is a tussle between territoriality and extraterritoriality as the lens through which to view these questions. The challenge is not to try and isolate nation-state laws as though we are living in 1925, but to identify the range of circumstances in which a particular assertion of one law over another is acceptable or unacceptable.

II. APPLICABLE LAW IN U.S. COURTS

With these systemic observations in mind, what of the private international law issues as U.S. courts would confront them? I will focus on questions of applicable law. In this regard, one must differentiate between how U.S. courts would determine the law applicable to a publicity rights violation under state law, on the one hand, and a claim under section 43(a) of the Lanham Act or a new federal statute (such as the proposed NO FAKES Act) on the other hand.[22]

A state publicity claim would be analyzed using one of several choice of law methods familiar to U.S. courts. Prevailing approaches vary widely. But they include contemporary interest analysis as well as the more traditional adoption of a single localization factor such as *lex loci protectionis*, *lex loci delicti*, or law of the plaintiff's domicile.[23] By contrast, the federal claims would likely turn at least in part on the application of the two-step test first articulated by the U.S. Supreme Court outside intellectual property law in *Morrison v. National Australia Bank*.[24]

A. STATE PUBLICITY RIGHTS CLAIM

One existing form of intellectual property most directly implicated by the distribution of deepfakes is state publicity law. Professor Kim Roosevelt in 1999 famously wrote that "[c]hoice of law is a mess."[25] Maybe it is. But choice of law in publicity cases is *really* a mess. To be sure, there are a good number of reported decisions. But there is nothing approaching a consensus (or even a basic method) that holds that case law together.

Despite that, there are a couple of observations one can make about the body of case law and the relatively light scholarly commentary.[26] First, the case law highlights a relatively greater role for the law of the domicile[27] than one would typically find in

22. *See* NO FAKES Act of 2025, S. 1367, 119th Cong. (2025); *see* Jennifer E. Rothman, *Reframing Deepfakes*, 49 COLUM. J.L. & ARTS 685 (2026).

23. *See* Symeon C. Symeonides, *The Need for a Third Conflicts Restatement (And a Proposal for Tort Conflicts)*, 75 IND. L.J. 437 (2000).

24. Morrison v. National Australia Bank Ltd., 561 U.S. 247, 255, 266 (2010).

25. Kermit Roosevelt III, *The Myth of Choice of Law: Rethinking Conflicts*, 97 MICH. L. REV. 2448, 2449 (1999).

26. For a helpful attempt to pull some of the case law together, see Mary LaFrance, *Choice of Law and the Right of Publicity: Rethinking the Domicile Rule,* 37 CARDOZO ARTS & ENT. L.J. 1, 6–19 (2019).

27. *See, e.g.*, Cairns v. Franklin Mint Co., 292 F.3d 1139, 1149 (9th Cir. 2002); Rogers v. Grimaldi, 875 F.2d 994, 1002 (2d Cir. 1989); Southeast Bank, N.A. v. Lawrence, 489 N.E.2d 744, 745 (N.Y. 1985); J. THOMAS MCCARTHY & ROGER E. SCHECHTER, RIGHTS OF PUBLICITY AND PRIVACY § 11:15 (2d ed. 2025).

transnational intellectual property disputes.[28] Several cases in U.S. courts have examined whether publicity right protections are available to foreign plaintiffs.[29] In some of those cases, courts answered that question by asking whether the plaintiff had such rights under the law of the place of his or her residence or domicile (at death, where relevant).[30]

This choice of connecting factor may have been consistent with trends in the Brussels jurisdictional rules after *eDate* and *Martinez*,[31] and indeed Italian choice of law rules under article 24, or article 95 of the pending French codification project.[32] This is arguably justified because of the personal nature of the image rights or publicity rights, which are understood conceptually as emanating from privacy concerns. So, it is tempting to see a shared transatlantic focus on the place of domicile in these types of cases, reflecting a common view of the types of governmental interest at stake in publicity cases.

But that would be a mistake. As a descriptive matter, the apparent emphasis on domicile might result from the types of issues that have received the attention of U.S. courts. In many of these domicile-grounded opinions, the particular question before the court was whether *post-mortem* publicity rights were available.[33] Courts in those cases might naturally seek to apply the law of a single state to the property of an estate and thus to elevate place of domicile as a localization rule that can achieve that goal.

---

28. The nationality of the author of a work, or the owner of a trademark, can be relevant to the country origin of a work. In turn this can affect the scope of protection of the international system. Despite that, the plaintiff's domicile might at first glance appear to be the third rail in transborder intellectual property litigation, because public international treaties require national treatment. *See* Paris Convention for the Protection of Industrial Property, Mar. 20, 1883, as last revised at Stockholm art. 22, July 14, 1967, 828 U.N.T.S. 305; Berne Convention for the Protection of Literary and Artistic Works (1971 Paris text) [hereinafter "Berne Convention"], art. 5, July 24, 1971, 1161 U.N.T.S. 3; Agreement on Trade-Related Aspects of Intellectual Property Rights art. 3, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization Annex 1C, LEGAL INSTRUMENTS—RESULTS OF THE URUGUAY ROUND VOL. 31, 33 I.L.M. 81 (1994) [hereinafter "TRIPS"]. Generally, that commitment imposes upon nation-states the obligation to treat foreign right holders at least as favorably as they treat local authors. Despite that, private international law analysis in the United States has managed to take nationality of the plaintiff into account both in regard to the applicable law and the application of *forum non conveniens*. This has conventionally been thought not to implicate the national treatment obligation. *See* Murray v. Brit. Broad. Corp., 81 F.3d 287, 291 (2d Cir. 1996); Creative Tech., Ltd. v. Aztech Sys. Pte., Ltd., 61 F.3d 696, 699 (9th Cir. 1995); Itar-Tass Russian News Agency v. Russian Kurier, Inc., 153 F.3d 82, 90–91 (2d Cir. 1998) (considering nationality as relevant to the law applicable to the question of copyright ownership). Recent developments in the World Trade Organization might, if followed, suggest a greater willingness to subject national rules of private international law to the discipline of public international law. *See* Award of the Arbitrators, China—Enforcement of Intellectual Property Rights, Arbitration Under Article 25 of the DSU, ¶ 4.72, WTO Doc. WT/DS611/ARB25 (July 21, 2025) (rejecting the arguments made by China that the TRIPS Agreement does not address issues of private international law).

29. *Cairns*, 292 F.3d at 1149.

30. *Id.*

31. Judgement for the Grand Chamber ¶¶ 1, 7, 13, Joined Cases C-509/09 & C-161/10, eDate Advert. GmbH v. X and Martinez v. MGN Ltd., 2011 E.C.R. 2011 I-10269.

32. *See* Treppoz, *supra* note 1, at 795-796.

33. *See, e.g.*, *Cairns*, 292 F.3d at 1146–47.

Thus, courts and commentators in turn have noted the centrality of domicile in publicity choice of law analyses.[34] It's a natural common law method to extrapolate from a series of similar instances. But it might distort principles, just as some conflicts scholars have argued that contemporary U.S. tort choice of law rules have been unduly shaped by working out whether guest statutes should preclude liability when New Yorkers crash their cars while driving neighbors to Canada.[35]

A fuller assessment of applicable law in publicity law would better embrace the more nuanced and contemporary approach to choice of law followed in most states of the United States. In the United States, publicity rights are clearly now seen as a property law concept.[36] The philosophical roots are complex and did indeed include privacy impulses and values tied to the person.[37] Over time, however, they have come to be seen as a conceptually distinct cause of action focused on commercial exploitation of identity. Publicity clearly has some of the characteristics of intellectual property and is understood by many as being subsumed within that category. If that were the case, we would not typically make the existence of a claim turn on the country of origin, in part because of national treatment commitments in public international treaties.[38]

Although the ALI Principles do not embrace a full governmental interest analysis, they clearly endorse an approach that broadens the range of relevant connecting factors. Thus, in section 301(2) the Principles suggest that the applicable law in a publicity case be "the law of each State in which direct and substantial damage results or is likely to result, irrespective of the State or States in which the act giving rise to the damage occurred."[39] In adopting this rule, the ALI assimilated publicity rights to "a noncontractual obligation arising out of an act of unfair competition."[40] The Reporters recognized that this departed from the approach in true privacy cases because "[t]he better view is to consider the right of publicity as an economic rather than a personal right, because its essence is to control exploitation."[41]

The ALI Reporters are surely correct that publicity rights now differ from "privacy rights [which] seek to prevent intrusion."[42] This is a persuasive assessment of the

34. *See* LaFrance, *supra* note 26, at 2; ALI PRINCIPLES, *supra* note 13, § 301, reporter's note 5 .

35. *See* Babcock v. Jackson, 12 N.Y.2d 473 (1963); Schultz v. Boy Scouts of Am., Inc., 65 N.Y.2d 189, 199 (1985) ("Although most of our major choice-of-law decisions after *Babcock* involved foreign guest statutes in actions for personal injuries, we have not so limited them, but have applied the *Babcock* reasoning to other tort issues as well."); Friedrich K. Juenger, *Choice of Law in Interstate Torts*, 118 U. PA. L. REV. 202 (1969) (critiquing what the obsession with guest statutes had wrought).

36. *Cf.* Itar-Tass Russian News Agency v. Russian Kurier, Inc., 153 F.3d 82, 90 (2d Cir. 1998) (drawing on modern choice of law rules applicable to property claims to develop for the first time a federal choice of law rule determining applicable law for issues of initial ownership of copyright).

37. *See* JENNIFER ROTHMAN, THE RIGHT OF PUBLICITY: PRIVACY REIMAGINED FOR A PUBLIC WORLD (2018) (offering the authoritative account).

38. *See* TRIPS, *supra* note 28, art. 3 (national treatment obligation); Berne Convention, *supra* note 28, art. 5(1) (national treatment obligation).

39. ALI PRINCIPLES, *supra* note 13, § 301(2).

40. *See id.*

41. *Id.* at cmt. e.

42. *Id.*

governmental interests that have driven the evolution of publicity rights in the United States over the last seventy-five years. But whether that is true of what drives efforts behind the use of publicity rights to preclude deepfakes is, however, a closer case. One could detect an array of governmental interests in these putative cases, including autonomy, dignity, market harm, deception, and the erosion of social fabrics.[43] Regulating these harms implicate a wide range of governmental interests.

Given this messy multivalence, it might well be that we also should consider a more general tort approach and apply the law of the most significant relationship, while recognizing that one of the dominant connecting factors is place of damage.[44] This would not entirely undermine the prospect of a single law being applied to an entire transborder dispute, and thus would retain some of the benefits of a country of origin (or domicile)_rule.

An interest analysis would also allow courts room to recognize the different range of values that might be presented by publicity cases and also would place less dispositive weight on identifying the place of damage. Notably, the ALI did not say anything further about where the "the place of damage" would be. The illustration in the ALI Principles assumes multiple places of damage, all where the act occurred, bringing it closer to endorsing the *lex loci protectionis* than supporting the law of a single place of localized reputation.[45]

B. TRADEMARK AND UNFAIR COMPETITION (FALSE ENDORSEMENT) CLAIMS

The other most likely claim that might be asserted under current U.S. law would be false endorsement under section 43(a) of the Lanham Act.[46] This claim has its substantive weaknesses, and the likely relief (a disclaimer) may cause it to sound more in transparency values. But many of the leading publicity rights cases have parallel section 43(a) claims.[47]

---

43. *See* Rothman, *supra* note 22.

44. *See* RESTATEMENT (SECOND) OF CONFLICT OF LAWS § 145(1) (A.L.I. 1971) ("The rights and liabilities of the parties with respect to an issue in tort are determined by the local law of the state which, with respect to that issue, has the most significant relationship to the occurrence and the parties under the principles stated in § 6."); *id.* at § 145(2) ("Contacts to be taken into account in applying the principles of § 6 to determine the law applicable to an issue include: (a) the place where the injury occurred, (b) the place where the conduct causing the injury occurred, (c) the domicil, residence, nationality, place of incorporation and place of business of the parties, and (d) the place where the relationship, if any, between the parties is centered.").

45. ALI PRINCIPLES, *supra note* 13, § 301(2), illustration 1. Oddly, in the "Rights of publicity" section, the ALI includes an illustration which assumes a court uses Section 301(1)(b), which applies the *lex loci protectionis*. *See id.* at § 403, cmt. c, illustration 2 (If "an advertisement for a UK company is broadcast in the United States using a picture of the United Kingdom's Prince Charming without his permission. Prince Charming sues for violation of his U.S. right of publicity. A U.S. court, following § 301(1)(b), applies U.S. law and awards damages . . . .").

46. *See* Rothman, *supra* note 22.

47. *See, e.g.*, White v. Samsung Elecs. Am., Inc., 971 F.2d 1395, 1396 (9th Cir. 1992); *see also* Waits v. Frito-Lay, Inc., 978 F.2d 1093, 1096 (9th Cir. 1992).

Publicity rights are not the only example of the unfair competition provision of the statute, section 43(a), being deployed to vindicate personality concerns. For example, in *Gilliam v. ABC*, the Second Circuit allowed Monty Python relief under section 43(a) for what was essentially a moral rights violation (another personality grounded claim).[48] Analogously, passing off (the core of English unfair competition law) was how the English court in *Fenty v. Arcadia* was able to fashion something approaching an image right.[49]

The territorial scope of federal legislation is these days determined under the Supreme Court's *Morrison* framework.[50] This created a two-step test. First, Congress is presumed to have intended statutes to apply only within the United States because Congress is generally concerned only with local conditions. Thus, unless the statute explicitly indicates a broader reach, it will have only domestic scope. Second, if a statute is not extraterritorial, a court must consider the "focus" of the statutory provision. If the focus is on activity (such as a particular act, injury, or effect) that occurred in the United States, then the application of the statute should not be regarded as extraterritorial. However, if the focus is on a particular activity that takes place abroad, then it is an impermissible extraterritorial application.[51] That test is unilateralist, interpreting the scope of the legislation within the confines of default assumptions including a renewed presumption against extraterritoriality that had slowly been weakened over the course of the twentieth century.[52]

The *Morrison* framework has in the past two years been applied by the U.S. Supreme Court in a Lanham Act case, *Abitron v. Hetronic*.[53] The effect of that case was substantially to curtail the extraterritorial application of the Lanham Act.[54] All Justices agreed that the Act did not apply extraterritorially.[55] Moreover, all Justices agreed that the statute could not reach foreign conduct such as sales of products abroad to foreign customers.[56] The plaintiff had sought to bring wholly foreign sales within the scope of the U.S. statute by asserting that the damage to its goodwill caused by such conduct occurred in the United States, where it was domiciled.

But beyond these questions the Court splintered quite badly. On the question of whether U.S. law reached conduct abroad that had effects on consumers in the United States, the Court nominally split 5–4 in favor of requiring some conduct in the United States. However, that majority was secured by the vote of Justice Jackson, whose opinion raised some doubt as to whether she might under different circumstances have

48. *See* Gilliam v. Am. Broad. Cos., 538 F.2d 14, 24 (2d Cir. 1976).
49. Fenty v. Arcadia Grp. [2015] EWCA (Civ) 3 (UK).
50. *See* Morrison v. National Australia Bank Ltd., 561 U.S. 247 (2010).
51. *See* Spanski Enters., Inc. v. Telewizja Polska, S.A., 883 F.3d 904, 909, 913 (D.C. Cir. 2018).
52. *See* William S. Dodge, *The New Presumption Against Extraterritoriality*, 133 HARV. L. REV. 1582 (2020).
53. Abitron Austria GmbH v. Hetronic Int'l, Inc., 600 U.S. 412, 412 (2023).
54. Steele v. Bulova Watch Co., 344 U.S. 280, 282, 286–87 (1952).
55. *See Abitron*, 600 U.S. at 412, 417, 424, 428.
56. *See Abitron*, 600 U.S. at 412, 437 (Sotomayor J., concurring in the judgment).

agreed with Justice Sotomayor's rival opinion.[57] Justice Sotomayor's opinion for four Justices would have allowed the U.S. statute to reach conduct abroad if it caused consumer confusion in the United States.[58]

Earlier non–Lanham Act case law clearly stated that "the focus of a statute [which had to occur in the U.S. for the statute to apply] was 'the object of its solicitude,' which can *include* the conduct it 'seeks to "regulate,"' *as well* as the parties and interests it "'seeks to 'protect' or 'vindicate.'"[59] But Justice Alito held that step two does not end with identifying statutory focus and instead courts must "as[k] whether *the conduct* relevant to that focus occurred in United States territory."[60]

Formally, Justice Jackson's joining of the Alito opinion means that a third step was added to *Morrison*'s two step test—a court must ask whether the "*conduct* relevant to [the focus of the statute] occurred in United States territory."[61] That test would on its face substantially limit the capacity to apply the Lanham Act to a deepfake created abroad that harmed the plaintiff within the United States.

However, Justice Jackson does acknowledge in footnote 2 of her concurring opinion, that uses on a server abroad might be within the scope of the statute.[62] She was clearly very much on the fence between the dueling opinions, and it does not require much judicial imagination to define the relevant conduct—use of the mark—as involving effects in the United States. This is explicitly what the WIPO Joint Recommendation on use on the internet did in 2001, a shift that has also been reflected in numerous jurisdictional and substantive assessments of trademark claims online.[63]

At the time the opinion was handed down, I thought that Justice Sotomayor would eventually prevail. But the lower courts have thus far stuck quite closely to Justice Alito's line. On remand in *Abitron* itself, for example, the Tenth Circuit reversed (as the Supreme Court plainly required) its earlier decision to apply the Lanham Act to foreign sales of goods that did not reach the United States but which diverted customers from the US producer (and thus harmed the US economy).[64] But the court also stressed that

57. *Id.* at 430–31 (Jackson, J., concurring)

58. *Id.* at 433 (Sotomayor, J., concurring in the judgment).

59. WesternGeco LLC v. ION Geophysical Corp., 585 U.S. 407, 414 (2018) (emphasis added).

60. *Abitron*, 600 U.S. at 418 (emphasis added).

61. *Id.*; see also *id.* at 430 (Jackson, J., concurring) ("[I]f the mark is not serving that function in domestic commerce, then the conduct Congress cared about is not occurring domestically, and these provisions' purely domestic sweep cannot touch that person."); *see also id.* at 439 (Sotomayor J., concurring in the judgment) ("[I]nstead of discerning the statute's focus and assessing whether that focus is found domestically, as the Court's precedents command, the majority now requires a third step: an assessment of whether the 'conduct relevant to the focus' occurred domestically, even when the focus of the statute is not conduct.").

62. *Abitron*, 600 U.S. at 432 n.2 (Jackson, J., concurring) ("[I]n the internet age, one could imagine a mark serving its critical source-identifying function in domestic commerce even absent the domestic physical presence of the items whose source it identifies.") (citing 5 J. MCCARTHY, TRADEMARKS AND UNFAIR COMPETITION § 29:56 (5th ed. Supp. 2023) ("The use of an infringing mark as part of an Internet site available for use in the United States may constitute an infringement of the mark in the United States.")).

63. *See* Joint Recommendation (WIPO), *supra* note 10.

64. Hetronic Int'l, Inc. v. Hetronic Ger. GmbH, 99 F.4th 1150 (10th Cir. 2024).

the Supreme Court opinion "counsels against our assigning significance to the ultimate destination of the infringing goods. . . . What matters is 'the location of the conduct relevant to the focus.'"[65] Thus, foreign sales of goods that ended up in the United States because of some later downstream sales by others could not render the foreign distributor liable under the Lanham Act. The court explained that "products bound for the United States but sold abroad cannot premise a Lanham Act claim without some domestic conduct tying the sales to an infringing use of the mark in U.S. commerce."[66] This conclusion was not altered by the fact that Abitron "took steps to facilitate [the] sales" in the United States—namely, by obtaining FCC licenses and hiring a U.S.-based distributor."[67] The only glimmer of hope that Justice Sotomayor's view would carry the day in the long-term was that the Tenth Circuit remanded to the district court to consider whether the plaintiff could establish domestic advertising, leaving it open for lower courts to explore the tension in Justice Jackson's footnote 2 by probing further where online advertising "occurs."

But it is a glimmer only. In a recent Ninth Circuit opinion, in *Doctor's Best, Inc. v. Nature's Way Products,* the Court blended the requirement of U.S. *conduct* for subject matter jurisdiction with the element of U.S. *effects* (namely, confusion) in the liability standard.[68] As a result, the case was summarily dismissed. A concurring judge went so far as to say that "Lanham Act only applies where there is a domestic use in commerce that would cause a domestic likelihood of confusion."[69] So, we have reached the stage where the application of the statute is not a contest between conduct and effects, but rather a demand for both.

This substantially weakens the ability of the Lanham Act to assist. And if a similar framework were to be applied to any future federal legislation such as the NO FAKES Act, it could make it quite toothless, which is something I hope drafters of any such legislation bear in mind. Some extraterritorial effects are inevitable in the deepfakes context. And as the drafters of the EU AI Act have shown, no one can reasonably be offended if the U.S. Congress wishes to enter that arena to protect important values.

65. *Id.* at 1166.
66. *Id.* at 1169.
67. *Id.*
68. Doctor's Best, Inc. v. Nature's Way Prods., LLC, 143 F.4th 1101, 1105, 1107–08 (9th Cir. 2025).
69. *Id.* at 1114 (Ikuta, J., concurring).

Deepfakes, Real Enforcement Challenges

David S. Louk*

* J.D., Yale Law School; Ph.D., Jurisprudence & Social Policy, UC Berkeley; Deputy City Attorney, San Francisco City Attorney's Office; Former Academic Fellow, Columbia Law School. The author wishes to thank Professor Jane Ginsburg, the faculty and staff of the Kernochan Center, and the student editors of the *Columbia Journal of Law & the Arts* for hosting the excellent symposium out of which this Article formed, as well as for their skilled and diligent edits to the Article. No non-public information was relied upon in researching and writing this Article. All views professed in this piece are this author's alone and do not reflect the position of any other individuals or institutions.

INTRODUCTION

The advent and rapid advancement of generative AI ("gen AI") technology over the past several years has been accompanied by a proliferation of AI-generated deepfake pornography, which depicts real people in intimate ways that they neither participated in, nor consented to. The ability to cheaply and easily generate realistic-looking, AI-generated audio, video, or images that convincingly mimic real individuals depicted in intimate contexts without their consent, known as non-consensual intimate depictions ("NCID"), presents a unique challenge in technology regulation. This Article focuses on a subset of those depictions, non-consensual intimate images (or "NCII").[1] In contrast to the contentious and heavily debated free-speech and intellectual property issues raised by artificial intelligence and deepfakes in other contexts, NCII features a

1. *See Nonconsensual Distribution of Intimate Images: What to Know*, FED. TRADE COMM'N: CONSUMER ADVICE (Nov. 2024), https://consumer.ftc.gov/articles/nonconsensual-distribution-intimate-images-what-know [https://web.archive.org/web/20260206180803/https://consumer.ftc.gov/articles/nonconsensual-distribution-intimate-images-what-know].

comparatively resolute moral and legal consensus. Indeed, NCII is one of the rare contemporary political issues where there exists bipartisan and near-universal agreement that the practice must be stopped. Despite this, effective enforcement mechanisms to curtail the practice have remained somewhat elusive; the gap between recognizing the harm and implementing effective remedies has yet to be closed.

This Article examines the technological underpinnings of AI-generated NCII deepfakes and the legal apparatuses that have sprung up recently to address it, catalogs the practical and legal obstacles to successful enforcement, and proposes strategic approaches for more effective regulation. While the rapid development and deployment of gen AI technology means that no perfect solution is likely to emerge overnight. Instead, state and national legislatures, cross-border government enforcement offices, and leading multinational technology companies will each need to play critical roles to mitigate the worst of the harms inflicted by NCII deepfakes. This need is also urgent, because the technology to mass produce deepfake nonconsensual intimate video content is rapidly developing, which will surely only multiply the problem.

I. DEFINING THE PROBLEM: NCII DEEPFAKES

Non-consensual intimate images convincingly depict real individuals in false but intimate contexts—and without their consent. To date, existing technology has primarily supported the easy and widespread dissemination of static images, and so this Article's focus is on NCII, though widespread audio and video depictions are sure to follow. NCII differs from the emergence of so-called "revenge porn" cases in the early 2010s, which typically featured *real* intimate content but that was distributed without the consent of the depicted.[2] By contrast, contemporary NCII deepfakes involve wholly fabricated depictions of real individuals who neither consented to their creation nor participated in any underlying intimate activity. Worse, an even more troubling subset of NCII involves Child Sexual Abuse Material (CSAM), which depict minors in sexual contexts, and which can be created from any available image of a clothed minor.[3]

A. OPEN SOURCE MODELS AND THE EROSION OF GUARDRAILS

Given widespread recognition of the social harms and illegality of NCII deepfakes, how have they proliferated so rapidly? At present, most major gen AI platforms, such as ChatGPT's Dall-E and Sora image- and video-generating tools, have implemented guardrails that seek to prevent NCII and CSAM generation.[4] Nevertheless, these

2. *See, e.g.*, Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 FLA. L. REV. 1251, 1257–61 (2017).

3. *What Is NCII?*, INHOPE (Feb. 17, 2023), https://web.archive.org/web/20250124021021/https://inhope.org/EN/articles/what-is-ncii.

4. *See, e.g.*, *Child Safety: Adopting Safety by Design Principles*, OPENAI (Apr. 23, 2024) https://openai.com/index/child-safety-adopting-sbd-principles/ [https://web.archive.org/web/20260210060944/https://openai.com/index/child-safety-adopting-sbd-

protections can be circumvented through prompt manipulation,[5] and numerous web forums are dedicated to guiding users in how to override such guardrails.[6] Indeed, Grok, xAI's gen AI image generating software, experienced widespread issues with NCII creation notwithstanding claims it was not designed to produce such content.[7] (More on this below.) What all closed-source models share in common, however, is the ability to implement new safeguards, alter input sanitation methods, and, in extreme circumstances, limit access to the model altogether.

More significant, however, are the systemic vulnerabilities presented by open-source models, a veritable Pandora's box that, once opened, cannot be closed down. For example, when Stability AI released a version of its Stable Diffusion model in 2022, version 1.5, the open-source text-to-image generator inadvertently contained training

principles/] ("strong guardrails and safety measures" in ChatGPT and DALL-E); *Sora System Card*, OPENAI (Dec. 9, 2024) https://openai.com/index/sora-system-card/ [https://web.archive.org/web/20260303044058/https://openai.com/index/sora-system-card/] ("multi-tiered moderation strategy" including classifiers and blocklists); *Building Safeguards for Claude*, ANTHROPIC (Aug. 12, 2025) https://www.anthropic.com/news/building-safeguards-for-claude [https://web.archive.org/web/20260206195902/https://www.anthropic.com/news/building-safeguards-for-claude] (CSAM detection through hash comparison); *Gemini for Safety Filtering and Content Moderation*, GOOGLE CLOUD (updated Mar. 5, 2026) https://docs.cloud.google.com/vertex-ai/generative-ai/docs/multimodal/gemini-for-filtering-and-moderation [https://web.archive.org/web/20260206195940/https://docs.cloud.google.com/vertex-ai/generative-ai/docs/multimodal/gemini-for-filtering-and-moderation] (CSAM safety filters).

5. *See, e.g.*, Emilia Napolano, *Sora: Inappropriate and Harmful Content Creation Easily Bypassed Through Simple Prompt Engineering*, ZENODO (Apr. 26, 2025), https://zenodo.org/records/15295087 [https://perma.cc/7T5G-QFZA]. Users are also able to use image-generating platforms to create "fetish content" of real individuals. Katie Notopoulos, *Sora Might Have a "Pervert" Problem on Its Hands*, BUS. INSIDER (Oct. 24, 2025), https://www.businessinsider.com/sora-video-openai-fetish-content-my-face-problem-2025-10 [https://web.archive.org/web/20260102103649/https://www.businessinsider.com/sora-video-openai-fetish-content-my-face-problem-2025-10].

6. Prompt manipulation techniques—commonly called "jailbreaking"—can bypass AI safety guardrails through various methods, including roleplay scenarios, encoded instructions, multi-turn desensitization, and contextual manipulation. *See* Yi Liu et al., *Jailbreaking ChatGPT via Prompt Engineering: An Empirical Study*, ARXIV (May 23, 2023) https://arxiv.org/abs/2305.13860 [https://web.archive.org/web/20260204204720/https://arxiv.org/abs/2305.13860] (empirical study of seventy-eight jailbreak prompts across ten distinct patterns including pretending, attention shifting, and privilege escalation); *see also* Danny Bradbury, *Researchers Break OpenAI Guardrails*, MALWAREBYTES (Oct. 13, 2025), https://www.malwarebytes.com/blog/news/2025/10/researchers-break-openai-guardrails [https://web.archive.org/web/20251109062710/https://www.malwarebytes.com/blog/news/2025/10/researchers-break-openai-guardrails] (describing "Policy Puppetry" technique that bypassed safety measures across ChatGPT, Claude, and Gemini by disguising prompts as configuration files); Victor Tangermann, *Researchers Find Easy Way to Jailbreak Every Major AI, from ChatGPT to Claude*, FUTURISM (Apr. 25, 2025), https://futurism.com/easy-jailbreak-every-major-ai-chatgpt [https://web.archive.org/web/20251125080933/https://futurism.com/easy-jailbreak-every-major-ai-chatgpt] (reporting HiddenLayer exploit bypassed safety guardrails across all major frontier AI models using prompt injection combined with roleplaying). Unfortunately, research indicates these attacks succeed approximately 20% of the time, requiring an average of forty-two seconds and as few as five interactions to bypass safety guardrails—some succeed in under four seconds. *What Is AI Jailbreaking?*, IRONSCALES (Nov. 2025), https://ironscales.com/glossary/what-is-ai-jailbreaking [https://web.archive.org/web/20251125161610/https://ironscales.com/glossary/what-is-ai-jailbreaking] (citing IBM Research 2024 data).

7. *See infra* section I.C.

data that included intimate and CSAM content.[8] Because open-source models can be freely downloaded and retrained on user-selected datasets, bad actors quickly repurposed this technology.[9] The result was a proliferation of "nudification" and "undressing" applications and websites starting in 2023,[10] which promoted their services with taglines like "[i]magine wasting time taking her out on dates, when you can just use [the website] to get her nudes."[11] These platforms leveraged retrained models specifically optimized for generating non-consensual intimate content.[12]

Thus, by September 2023, an estimated 24 million users per month visited nudification websites—a figure likely significantly higher today.[13] Analysis of deepfake content reveals that approximately 96–99% of deepfake pornography targets women,[14] and a 2024 survey by Thorn found that among 13–20 year-olds, one in eight personally know someone who has been a victim of an NCII deepfake.[15] Additionally, the Center

8. David Evan Harris & Dave Willner, *Was an AI Image Generator Taken Down for Making Child Porn?*, IEEE SPECTRUM (Aug. 30, 2024), https://spectrum.ieee.org/stable-diffusion [https://web.archive.org/web/20251129101851/https://spectrum.ieee.org/stable-diffusion].

9. William Hawkins, Chris Russell & Brent Mittelstadt, *Deepfakes on Demand: The Rise of Accessible Non-Consensual Deepfake Image Generators*, FACCT '25: PROCS. OF THE 2025 ACM CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1, 1 (2025).

10. *See id.* at 1, 2 (finding that "both Stable Diffusion and Flux models are used for the creation of deepfake models, with 96% of these targeting women," and identifying over 34,000 downloadable deepfake model variants intended to generate images of identifiable individuals); Kyle Wiggers, *Deepfakes for All: Uncensored AI Art Model Prompts Ethics Questions*, TECHCRUCNCH (Aug. 26, 2022), https://techcrunch.com/2022/08/24/deepfakes-for-all-uncensored-ai-art-model-prompts-ethics-questions/
[https://web.archive.org/web/20251113085158/https://techcrunch.com/2022/08/24/deepfakes-for-all-uncensored-ai-art-model-prompts-ethics-questions/] ("Creative and malicious users can abuse the capabilities [of Stable Diffusion] to generate subjectively objectionable content at scale, using minimal resources to run inference —which is cheaper than training the entire model— and then publish them in venues like 4chan.").

11. First Amended Complaint ¶ 6, People v. Sol Ecom, Inc., No. CGC-24-617237 (Cal. Super. Ct. filed Mar. 10, 2025).

12. *See* Santiago Lakatos, *A Revealing Picture*, GRAPHIKA (Dec. 8, 2023), https://graphika.com/reports/a-revealing-picture
[https://web.archive.org/web/20260201082539/https://graphika.com/reports/a-revealing-picture].

13. *Id.*; *see also* Margi Murphy, *"Nudify" Apps That Use AI to "Undress" Women in Photos Are Soaring in Popularity*, TIME (Dec. 9, 2023), https://time.com/6344068/nudify-apps-undress-photos-women-artificial-intelligence/ [https://web.archive.org/web/20251223195706/https://time.com/6344068/nudify-apps-undress-photos-women-artificial-intelligence/].

14. Hailey Reissman, *What Is Deepfake Porn and Why Is It Thriving in the Age of AI?*, ANNENBERG SCH. COMMC'N UNIV. PA. (July 13, 2023), https://www.asc.upenn.edu/news-events/news/what-deepfake-porn-and-why-it-thriving-age-ai
[https://web.archive.org/web/20251128224820/https://www.asc.upenn.edu/news-events/news/what-deepfake-porn-and-why-it-thriving-age-ai]; Manjeevan Singh Seera & Ridoan Karim, *Digital Child Abuse: Deepfakes and the Rising Danger of AI-Generated Exploitation*, LENS (MONASH UNIV.) (Feb. 25, 2025), https://lens.monash.edu/@politics-society/2025/02/25/1387341/digital-child-abuse-deepfakes-and-the-rising-danger-of-ai-generated-exploitation
[https://web.archive.org/web/20251023085721/https://lens.monash.edu/@politics-society/2025/02/25/1387341/digital-child-abuse-deepfakes-and-the-rising-danger-of-ai-generated-exploitation].

15. Olina Banerji, *More Teens Than You Think Have Been "Deepfake" Targets*, EDUC. WEEK (Mar. 3, 2025), https://www.edweek.org/technology/more-teens-than-you-think-have-been-deepfake-

for Democracy & Technology's 2024 survey of high school students found that 15% (representing approximately 2.3 million students) had heard of an NCII deepfake depicting someone at their school during the 2023–24 school year.[16] (These numbers are almost certainly higher today.) Celebrity NCII deepfake content has proliferated through dedicated forums where users solicit custom content for cryptocurrency payments, creating a shadow economy around non-consensual sexual imagery.[17]

B. TECHNICAL EVASION OF DETECTION SYSTEMS

In the face of this profound diffusion of morally troublesome content, the easiest solution would be to implement sound guardrails on the technology itself, but as noted, no guardrails can guarantee prevention of the production of NCII. Closed-source image-generating models like Grok—which permit generation of "spicy" content—have had documented failures in preventing NCII generation that has recently resulted in several prominent lawsuits from victims, even as the platform disclaimed any intent to allow the creation of such content. Open-source platforms fare even worse: Once an open-source model has been released that can be trained to generate deepfake NCII content, in the absence of a kill-switch embedded by the model's creator, there is no way to "shut off" the technology. In such circumstances, the next best approach would be to develop technologies to identify and remove deepfake NCII content, but current detection and attribution technologies remain inadequate. Studies have shown that human detection of deepfake content generally is barely above chance, with overall accuracy of only 55.54%, and odds ratios indicating detection accuracy as low as 39%.[18] In theory, watermarks embedded in deepfake content can provide notice to viewers that content is AI-generated, potentially mitigating deception.[19] However, while some platforms embed watermarks or metadata in generated content, tools can readily remove these markers: Indeed, users of deepfake NCII creation tools will pay premium

---

targets/2025/03 [https://web.archive.org/web/20260123083556/https://www.edweek.org/technology/more-teens-than-you-think-have-been-deepfake-targets/2025/03].

16. Elizabeth Laird, Maddy Dwyer & Kristin Woelfel, *Report—In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K-12 Schools*, CTR. FOR DEMOCRACY & TECH. (Sep. 26, 2024), https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/ [https://web.archive.org/web/20260126002252/https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/]; Kara Arundel, *Schools Lack Supports for Victims of Sexually Explicit Deepfake and Real Images*, K-12 DIVE (Sep. 26, 2024), https://www.k12dive.com/news/schools-deepfake-images-student-supports/728107/ [https://web.archive.org/web/20260126002252/https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/].

17. Lakatos, *supra* note 12.

18. Alexander Diel et al., *Human Performance in Detecting Deepfakes: A Systematic Review and Meta-Analysis of 56 Papers*, 16 COMPUT. IN HUM. BEHAV. REPS. 1, 4, 6 (2024).

19. Nicola Henry, *AI "Nudify" Sites Are Being Sued for Victimising People. How Can We Battle Deepfake Abuse?*, CONVERSATION (Aug. 21, 2024), https://theconversation.com/ai-nudify-sites-are-being-sued-for-victimising-people-how-can-we-battle-deepfake-abuse-237043 [https://web.archive.org/web/20251204115841/https://theconversation.com/ai-nudify-sites-are-being-sued-for-victimising-people-how-can-we-battle-deepfake-abuse-237043].

fees specifically to remove watermarks that identify images as fake, and watermarks can also be easily cropped using simple graphic editing software.[20] For example, in one survey of such apps, seven of twenty analyzed AI nudification applications watermarked their output images and sold removal of those watermarks at the highest subscription tier, suggesting that watermark removal is a built-in feature of the NCII creation ecosystem.[21]

Gen AI content can itself be detected by AI software, though, to date, with limited efficacy. Current deepfake detection technologies have limited effectiveness in real-world scenarios, with accuracy reduced when lighting conditions, facial expressions, or video quality differ from training data, and future advances in deepfake generation are expected to eliminate current detection hallmarks.[22] Metadata, though more difficult to remove, offers no obvious warning to consumers who view the images.[23] Neither approach provides sufficient fail-safe protection for victims. While transparency practices like watermarking and labeling may aid in accountability and content moderation, doing so does not render noticeably "fake" content unharmful; even noticeably fake NCII can still result in mental and physical harm, reputational damage, and financial costs to the victims depicted.[24]

C. DISTRIBUTION

A third challenge with preventing the spread of NCII deepfakes is the speed and scope with which they can be almost instantaneously disseminated worldwide. As discussed below, the legal distinctions between the websites and apps that generate the images, and the platforms (like Facebook, Snapchat, and X) that host them, have generally limited accountability against the platforms as hosts.

One rare instance where this has not been the case is Grok, the xAI gen AI tool that is integrated directly into X.com's social media website. Grok has recently faced widespread media attention—and subsequent lawsuits—after journalists, regulators, and advocacy groups documented in early 2026 that Grok could be used to generate deepfake NCII by taking photos of real individuals and "undressing" them

20. Marco Viola & Cristina Voto, *Designed to Abuse? Deepfakes and the Non-Consensual Diffusion of Intimate Images*, 201 SYNTHESE 1, 9 (Jan. 13, 2023).

21. Cassidy Gibson et al., *Analyzing the AI Nudification Application Ecosystem*, SEC '25: PROCS. OF THE 34TH USENIX CONF. ON SEC. SYMP. 1, 11 (Nov. 14, 2024).

22. *Science & Tech Spotlight: Combating Deepfakes*, U.S. GOV'T ACCOUNTABILITY OFF. (Mar. 11, 2024), https://www.gao.gov/products/gao-24-107292 [https://web.archive.org/web/20260129175027/https://www.gao.gov/products/gao-24-107292].

23. *See* Michelle L. Ding & Harini Suresh, *The Malicious Technical Ecosystem: Exposing Limitations in Technical Governance of AI-Generated Non-Consensual Intimate Images of Adults*, 2025 CONF. ON HUM. FACTORS IN COMPUTING SYS. SOCIOTECH. AI GOVERNANCE WORKSHOP (Apr. 24, 2025); Barry Collins, *AI or Not? How to Detect if an Image Is AI-Generated*, FORBES (Oct. 14, 2023), https://www.forbes.com/sites/barrycollins/2023/10/14/ai-or-not-how-to-detect-if-an-image-is-ai-generated/ [https://web.archive.org/web/20260401162430/https://www.forbes.com/sites/barrycollins/2023/10/14/ai-or-not-how-to-detect-if-an-image-is-ai-generated/].

24. Ding & Suresh, *supra* note 23.

without consent. Targets included adult women, public figures, and in some cases minors or apparent minors.[25]

What set Grok apart from the cascade of other nudifying websites that have proliferated in recent years is that it is not simply a gen AI model that can, perhaps inadvertently, generate deepfake NCII for an individual user; rather, Grok is integrated with X in a way that lets users easily feed in platform images and generate altered versions at scale, then circulate them on the same network. Grok was thus described as a "one-click harassment machine,"[26] with California Attorney General Rob Bonta accusing xAI of appearing to facilitate the large-scale production of deepfake NCII used to harass women and girls across the internet, including on X.[27]

In response to public backlash in mid-January 2026, *Wired* reported that after days of outrage, X had started limiting Grok image generation on X to paying subscribers, but that this did not solve the problem: Sexualized "undressing" images were still being produced, and the standalone Grok app and website could still generate harmful content.[28] Critics called that change inadequate insofar as it only reduced access on one surface while leaving abuse possible elsewhere—in their view, putting abuse behind a paywall rather than stopping it. Later that month, Attorney General Bonta sent xAI a cease-and-desist letter demanding that it stop the creation and distribution of deepfake NCII and child sexual abuse material, citing possible violations of California civil and criminal law.[29] As of the publication of this article, the Attorney General had not yet announced the results of this investigation.

In addition to the unusually integrated combination of gen AI model and social media platform, another key source of the problem was the gap between xAI's written rules and Grok's actual behavior—another theme discussed below. xAI's Acceptable Use Policy already said users must not violate statutory privacy rights, depict people's likenesses "in a pornographic manner," or sexualize children.[30] Yet despite such written prohibitions, Grok was still reported to be generating exactly that sort of material in practice, and appeared to lack sufficient guardrails to prevent it.

---

25. *See* Press Release, Rob Bonta, Att'y General (Cal.), Attorney General Bonta Launches Investigation into xAI, Grok over Undressed, Sexual AI Images of Women and Children (Jan. 14, 2026), https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-investigation-xai-grok-over-undressed-sexual-ai [https://web.archive.org/web/20260331030724/https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-investigation-xai-grok-over-undressed-sexual-ai].

26. Nilly Patel, *Why Nobody's Stopping Grok*, THE VERGE (Jan. 22, 2026), https://www.theverge.com/podcast/865275/grok-deepfake-undressing-elon-musk-content-moderation [https://web.archive.org/web/20260326132124/https://www.theverge.com/podcast/865275/grok-deepfake-undressing-elon-musk-content-moderation].

27. *See* Press Release, Bonta, *supra* note 25.

28. Matt Burgess, *X Didn't Fix Grok's "Undressing" Problem. It Just Makes People Pay for It*, WIRED (Jan. 9, 2026), https://www.wired.com/story/x-didnt-fix-groks-undressing-problem-it-just-makes-people-pay-for-it/ [https://web.archive.org/web/20260319224509/https://www.wired.com/story/x-didnt-fix-groks-undressing-problem-it-just-makes-people-pay-for-it/].

29. *See* Press Release, Bonta, *supra* note 25.

30. *xAI Acceptable Use Policy*, XAI (effective Jan. 2, 2025), https://x.ai/legal/acceptable-use-policy [https://web.archive.org/web/20260401152721/https://x.ai/legal/acceptable-use-policy].

The controversy with Grok and xAI thus encapsulates many of the challenges with deepfake NCII: the technological hurdles to prevent the creation of any such images, the speed with which they can be distributed, and the challenges with identifying the individuals who created the images (and, Grok notwithstanding, the websites that generated them).

II. THE EXISTING LEGAL FRAMEWORK

In response to the rise of NCII deepfakes, a broad, bipartisan consensus quickly emerged that legal, not just technological, solutions were necessary to address this problem. To date, a patchwork of state and federal laws—as well as overseas laws and regulations—seeks to address the problem.

A. FEDERAL PROHIBITIONS AND SECTION 230

Given the widespread emergence of NCII deepfakes and the largely bipartisan reaction against them, it should not be surprising that even a historically unproductive[31] Congress has recently enacted federal legislation to address aspects of the NCII deepfake problem. Until recently, while multiple federal laws have addressed *aspects* of deepfake NCII content, they did so by reaching deepfake NCII content only incidentally. Moreover, because none of these laws was designed to specifically address the problem, most focus largely on individual content creators or distributors. Such laws include, for example, longstanding state and federal prohibitions on the possession, distribution, and creation of CSAM, even of content not depicting real children.[32] Another federal law, enacted in 2016, prohibits non-consensual intimate disclosure, but it was originally designed to target so-called "revenge pornography," and was enacted against a backdrop where there was an expected relationship between the victim and the perpetrator.[33] In addition, general obscenity prohibitions in the U.S. Code have also provided additional grounds for prosecution of the *distribution* of AI-generated CSAM—if not the possession.[34]

31. Minho Kim & Ashley Wu, *How the House Slumped to Historic Lows of Productivity in 2025*, N.Y. TIMES (Jan. 16, 2026), https://www.nytimes.com/interactive/2026/01/17/us/politics/house-republicans-majority-productivity.html.

32. At the federal level, CSAM is criminalized under multiple statutes. 18 U.S.C. § 2251 prohibits sexual exploitation of children and production of CSAM, with mandatory minimum sentences of fifteen to thirty years for first-time offenders. Sections 2252 and 2252A criminalize transportation, distribution, receipt, and possession of CSAM, with mandatory minimums of five years for receipt and distribution offenses, and up to ten years for simple possession. 18 U.S.C. §§ 2252, 2252A. These statutes apply regardless of whether actual children were depicted.

33. See Consolidated Appropriations Act, 2022, Pub. L. No. 117-103, 136 Stat. 49, which amended 15 U.S.C. § 6851.

34. *See* Rianna Pfefferkorn, *Court Rules That Constitution Protects Private Possession of AI Generated CSAM*, TECH POLICY PRESS (Mar. 20, 2025), https://www.techpolicy.press/court-rules-that-constitution-protects-private-possession-of-aigenerated-csam/ [https://web.archive.org/web/20260306114005/https://www.techpolicy.press/court-rules-that-constitution-protects-private-possession-of-aigenerated-csam/] (discussing U.S. v. Anderegg, Case No. 24-

Moreover, with the exception of CSAM, section 230 of the Communications Decency Act has also, until recently, presented a significant obstacle to platform accountability for deepfake NCII.[35] Under section 230(c)(1), online platforms are not treated as the "publisher or speaker" of information provided by third-party users, effectively immunizing them from liability for user-generated content.[36] Courts have interpreted this provision broadly since the landmark 1997 decision in *Zeran v. America Online, Inc.*, which held that platforms cannot be held liable for third-party content even when they have knowledge of its illegal nature.[37] This sweeping immunity has historically prevented victims of deepfake NCII from pursuing claims against the platforms that host and distribute such images, as platforms successfully invoke section 230 to obtain early dismissal of suits that would treat them as publishers of harmful content.[38]

Given the widespread, bipartisan consensus that deepfake NCII content must be addressed, section 230's broad grant of immunity has recently been reconsidered by Congress. One early result is the]landmark federal law addressing deepfake NCII content, the Take It Down Act, which was signed into law on May 19, 2025, and extends liability beyond individual creators to platforms hosting, generating, or disseminating such content.[39] The Act criminalizes the knowing publication of NCII (both authentic and AI-generated) with penalties of up to two years imprisonment for content involving adults and up to three years for content involving minors.[40] Critically, it also requires covered platforms to establish notice-and-removal mechanisms and to remove reported NCII within forty-eight hours of receiving valid requests.[41] Enforcement authority resides with the Federal Trade Commission, which may treat violations as unfair or deceptive trade practices.[42]

---

cr-50-jdp (W.D. Wisc. Feb. 13, 2025) and noting that the court distinguished between the defendant's possession and distribution of virtual CSAM and denying defendant's motion to dismiss the *distribution* count).

35. 47 U.S.C. § 230 (2018).

36. *Id.* § 230(c)(1) ("No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.").

37. Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997) (holding that section 230 creates "a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service").

38. *See, e.g.*, Kaitlin O'Donnell, *Have We No Decency? Section 230 and the Liability of Social Media Companies for Deepfake Videos*, 2021 U. ILL. L. REV. 701, 704 (2021) ("The limitations of federal solutions are due, in part, to Section 230 of the Communication Decency Act which provides immunity to social media companies for content posted on their sites.").

39. Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act (TAKE IT DOWN Act), Pub. L. No. 119-12, 139 Stat. 55 (2025) (codified at 47 U.S.C. § 223) [hereinafter the "Take It Down Act"].

40. *Id.* § 2(a).

41. *Id.* § 3(a)(3).

42. *Id.* § 3(b); *see also* Billee Elliott McAuliffe & Clare H. Nowogrocki, *New Federal TAKE IT DOWN Act Gives Victims an Avenue for the Removal of Deepfake Images*, LEXOLOGY (May 19, 2025), https://www.lexology.com/library/detail.aspx?g=2fed5f0c-e9fa-4a45-ba3c-075866c92f49 [https://web.archive.org/web/20260207034708/https://www.lexology.com/library/detail.aspx?g=2fed5f0c-e9fa-4a45-ba3c-075866c92f49%2A%2A].

While the Take It Down Act's notice-and-takedown mechanisms for NCII represents a seeming carve-out from section 230 immunity,[43] the precise contours of the Take It Down Act's carveout for section 230 immunity remains unresolved, with commentators advancing mixed views about whether (and how) the Act overrides traditional platform immunity. Perhaps the majority view is that the Act eliminates section 230 protection, a position supported by a fair reading of its text. The Act is codified at section 223 of the Communications Act, which section 230(e)(1) explicitly exempts from immunity coverage.[44] Under this view, the Take It Down Act amends section 223 to impose notice-and-takedown obligations on platforms. Since section 230 immunity does not extend to section 223 offenses , platforms cannot invoke section 230 to defend against FTC enforcement actions for failing to comply with takedown requirements.[45] Moreover, proponents of this position argue that the Act "contradicts the basic immunity" of section 230(c)(1) by treating platforms *as publishers or speakers* when it imposes liability for their decisions regarding hosting and removing third-party content—precisely what section 230 was designed to prevent.[46] The Act creates asymmetric liability by providing a safe harbor only for good-faith removals, not for refusing removal requests, thereby incentivizing platforms to remove content without investigation to avoid penalties that can exceed $50,000 per violation.[47]

However, some questions remain as to whether section 230 immunity survives in circumstances outside FTC enforcement actions. This interpretation suggests that

43. *See* Jeffrey D. Neuburger & Jonathan Mollod, *Take It Down Act Signed into Law, Offering Tools to Fight Non-Consensual Intimate Images and Creating a New Image Takedown Mechanism*, PROSKAUER: NEW MEDIA & TECH. L. BLOG (May 29, 2025), https://newmedialaw.proskauer.com/2025/05/29/take-it-down-act-signed-into-law-offering-tools-to-fight-non-consensual-intimate-images-and-creating-a-new-image-takedown-mechanism/
[https://web.archive.org/web/20260207035255/https://newmedialaw.proskauer.com/2025/05/29/take-it-down-act-signed-into-law-offering-tools-to-fight-non-consensual-intimate-images-and-creating-a-new-image-takedown-mechanism/].

44. 47 U.S.C. §§ 230(e)(1), (e)(3) (2018) ("Nothing in this section shall be construed to impair the enforcement of . . . any . . . Federal criminal statute . . . or any . . . State law that is consistent with this section"); Take It Down Act § 2 (to be codified at 47 U.S.C. § 223(h)).

45. *See e.g.*, Neuburger & Mollod, *supra* note 43, ("Since the Take It Down Act states that it will be codified at section 223 of the Communications Act of 1934 (i.e., 47 U.S.C. 223(h)), it appears that platforms would not enjoy CDA protection from FTC civil enforcement actions based on the agency's authority to enforce the Act's requirements that covered platforms 'reasonably comply' with the new Take It Down Act notice-and-takedown obligations.").

46. *See, e.g.*, Thomas J. Cunningham & Michael J. McMorrow, *Platforms Face Section 230 Shift From Take It Down Act*, TROUTMAN PEPPER (June 9, 2025), https://www.troutman.com/insights/platforms-face-section-230-shift-from-take-it-down-act/
[http://web.archive.org/web/20260207041556/https://www.troutman.com/insights/platforms-face-section-230-shift-from-take-it-down-act/] ("[T]he act contradicts the basic immunity provided by [section 230] . . . , that '[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.' The act treats the providers or users of interactive computer services in exactly that way, by imposing liability on them as the publisher or speaker of 'information provided by another information content provider.'").

47. *Id.* (noting that failure to comply "shall be treated as a violation of a rule" under the FTC Act, with penalties currently at $53,088 per violation, and that the Act "provides no safe harbor . . . for rejecting or refusing to honor a request for removal"); *see also* Take It Down Act § 3(b)(2) .

while platforms likely cannot invoke section 230 against FTC civil enforcement for failure to implement compliant takedown procedures, they *would* retain immunity in other significant respects. First, it is arguable whether platforms maintain section 230 protection from claims related to hosting or publishing NCII that has not been the subject of a prior, valid Take It Down Act notice, since the Act's removal requirements are triggered only upon receiving a compliant takedown request.[48] Second, it may be the case that platforms could assert section 230 immunity against private lawsuits by individuals (as opposed to FTC enforcement) alleging harm from a platform's failure to comply with takedown notices, particularly since the Act creates no private right of action and platforms could argue such suits impermissibly treat them as publishers of third-party content.[49] Third, the Good Samaritan provision of section 230(c)(2) arguably shields platforms from liability for proactively filtering or removing NCII, which would be supplemented by the Act's own safe harbor for good-faith removals.[50] Under this narrower view, the Take It Down Act represents not a wholesale elimination of section 230 immunity but rather a targeted carve-out limited to federal enforcement of notice-and-takedown compliance obligations. Such a view is arguably further supported by the existence of additional proposed legislation that would more clearly strip all section 230 immunity from platforms, such as the Intimate Privacy Protection Act, which would condition section 230 immunity on platforms implementing a "duty of care" to prevent and remove deepfake NCII.[51] Because the Take It Down Act's takedown provisions do not go into effect until May 2026, the precise legal contours of the relationship between the Act and section 230 remain to be seen as of the publication of this article.

B. STATE-LEVEL RESPONSES

Given the severity of the problem posed by NCII deepfakes—and the broad bipartisan consensus against the practice—all fifty states have enacted some version of a non-consensual intimate image law, and at least forty-five states have subsequently amended or added specific prohibitions to target NCII deepfakes.[52] The broad,

48. Neuburger & Mollod, *supra* note 43 ("[T]he Act's requirements for removal of NCII by platforms would not be implicated without a valid removal request.").

49. *Id.* ("Similarly, a platform could make a strong argument that it retains CDA immunity from any claims brought by an individual (rather than the FTC) for failing to reasonably comply with a Take It Down Act notice.").

50. *Id.*

51. Press Release, Rep. Jake Auchincloss, Auchincloss Introduces Bipartisan Bill to Tackle Rise in Non-Consensual Deepfakes on Social Media Platforms (July 30, 2025), https://auchincloss.house.gov/media/press-releases/release-auchincloss-introduces-bipartisan-bill-to-tackle-rise-in-non-consensualdeepfakes-on-social-media-platforms [http://web.archive.org/web/20260207042913/https://auchincloss.house.gov/media/press-releases/release-auchincloss-introduces-bipartisan-bill-to-tackle-rise-in-non-consensualdeepfakes-on-social-media-platforms].

52. *See* Kaylee Williams, *Minors Are on the Frontlines of the Sexual Deepfake Epidemic—Here's Why That's a Problem*, TECH POL'Y PRESS (Oct. 10, 2024), https://www.techpolicy.press/minors-are-on-the-frontlines-of-the-sexual-deepfake-epidemic-heres-why-thats-a-problem/

bipartisan legislative action is reflective of widespread agreement that the problem demands intervention.

California's approach is illustrative of the evolving state-level response. Assembly Bill 602, enacted in 2019, created a civil cause of action for non-consensual image distribution (originally targeting "revenge pornography") and provided for disgorgement and up to $150,000 in statutory damages.[53] The statute applies to persons who create and intentionally disclose sexually explicit material when they know or reasonably should know the depicted individual did not consent, or who intentionally disclose such material they did not create knowing the depicted individual did not consent.[54] In 2025, the California legislature amended this statute to expressly enumerate liability for nudification websites, extending responsibility to platforms facilitating content creation and establishing specific statutory enforcement provisions for public prosecutors, including the attorney general and city and county attorneys.[55] Thus, just as the Federal Trade Commission (FTC) has been delegated primary enforcement authority over the Take It Down Act, California law anticipates the attorney general and city and county attorneys will have a central role to play in enforcement at the state level.

Finally, it is worth noting that the abovementioned state and federal prohibitions on NCII deepfakes also exist within a broader ecosystem of identity rights laws that also apply to NCII deepfakes. While not all states denominate these laws as "right of publicity" statutes, many have expressly adopted common-law privacy rights to protect against unauthorized appropriation of identity or recognized a "right of publicity" either by common law or by statute.[56] These laws have traditionally protected against unauthorized commercial exploitation of a person's name, image, and likeness (as well as, increasingly, voice), though their scope varies significantly from jurisdiction to jurisdiction. Additionally, federal and state trademark laws, unfair competition statutes, and consumer protection laws also restrict unauthorized uses of identity that

---

[http://web.archive.org/web/20260207043301/https://www.techpolicy.press/minors-are-on-the-frontlines-of-the-sexual-deepfake-epidemic-heres-why-thats-a-problem/]; *Revenge Porn Laws: State by State*, C.A. GOLDBERG, https://www.cagoldberglaw.com/resources/states-with-revenge-porn-laws/ [https://web.archive.org/web/20260226051618/https://www.cagoldberglaw.com/resources/states-with-revenge-porn-laws/] (last visited Apr. 1, 2026); *Tracker: State Legislation on Intimate Deepfakes*, Pub. Citizen (updated Mar. 31, 2026), https://www.citizen.org/article/tracker-intimate-deepfakes-state-legislation/ [https://web.archive.org/web/20260310101025/https://www.citizen.org/article/tracker-intimate-deepfakes-state-legislation/].

53. A.B. 602, 2021–22 Leg., Reg. Sess. (Cal. 2021), codified at CAL. CIV. CODE § 1708.86 (West 2021); Douglas E. Mirell & Joshua Geller, *AB 602 and AB 730: Curbing "Deepfakes" in Pornography and Elections*, DAILY J. (Jan. 8, 2020), https://www.dailyjournal.com/articles/355794-ab-602-and-ab-730-curbing-deepfakes-in-pornography-and-elections [http://web.archive.org/web/20260207044110/https://www.dailyjournal.com/articles/355794-ab-602-and-ab-730-curbing-deepfakes-in-pornography-and-elections].

54. CAL. CIV. CODE § 1708.86(b) (West 2025).

55. CAL. CIV. CODE § 1708.86 (West 2025) (as amended).

56. JENNIFER E. ROTHMAN, UNIFORM LAW COMMISSION PROTECTION OF NAME, IMAGE, [VOICE], AND LIKENESS STUDY COMMITTEE: REPORTER'S WELCOME MEMO 10–11 (2025); Jennifer E. Rothman & Robert Post, *The First Amendment and the Right(s) of Publicity*, 130 Yale L.J. 86, 94-95, 88 (2020).

create confusion or falsely indicates an endorsement.[57] As discussed below, these longstanding legal rights can also serve as a tool for enforcement against NCII deepfake content creation and distribution.[58]

III. PRACTICAL ENFORCEMENT CHALLENGES

A. ATTRIBUTION DIFFICULTIES

Identifying whether NCII content is a "deepfake" is challenging enough, but tracing content creation presents even greater difficulties.[59] Identifying both the individual prompting image generation and the specific application or website used proves exceptionally challenging.[60] Users typically operate behind VPNs or burner accounts, uploading content that rapidly proliferates across multiple platforms.[61] Within minutes of initial posting—particularly for celebrity content—attribution becomes functionally impossible as thousands of users redistribute images. For example, AI-generated sexually explicit images of musician and global celebrity Taylor Swift "gained over 45 million views, along with hundreds of thousands of likes, bookmarks, and reposts over a seventeen-hour period" before being taken down.[62] This viral spread outpaces effective legal response, creating a "whack-a-mole" problem where content removal efforts fail to contain dissemination.[63] (The lawsuits against Grok and xAI prove a notable exception, but Grok is also the rare image generator tied to a social media platform that reposts those same images, making attribution unusually straightforward.)[64]

57. *See* Jennifer E. Rothman, *Navigating the Identity Thicket: Trademark's Lost Theory of Personality, the Right of Publicity, and Preemption*, 135 HARV. L. REV. 1271, 1278–79 (2022).

58. *See infra* Part VI.A.

59. *See* Irene Amerini et al., *Deepfake Media Forensics: Status and Future Challenges*, 11 J. IMAGING 1, 29–30 (2025).

60. Gueltom Bendiab et al., *Deepfakes in Digital Media Forensics: Generation, AI-Based Detection and Challenges*, 88 J. INFO. SEC. & APPLICATIONS 1 (2025).

61. Catherine Han et al., *Characterizing the MrDeepFakes Sexual Deepfake Marketplace, in* SEC '25: PROCS. OF THE 34TH USENIX CONF. ON SEC. SYMP. 5169 (2025).

62. Halle Nelson, *Taylor Swift and the Dangers of Deepfake Pornography*, NAT'L SEXUAL VIOLENCE RES. CTR. (Feb. 7, 2024), https://www.nsvrc.org/blogs/feminism/taylor-swift-and-dangers-deepfake-pornography [https://web.archive.org/web/20260211200149/https://www.nsvrc.org/blog_post/taylor-swift-and-dangers-deepfake-pornography/].

63. Kat Tenbarge, *Nude Deepfakes of Taylor Swift Went Viral on X, Evading Moderation and Sparking Outrage*, NBC NEWS (Jan. 25, 2024), https://www.nbcnews.com/tech/misinformation/taylor-swift-nude-deepfake-goes-viral-x-platform-rules-rcna135669 [https://web.archive.org/web/20260211200554/https://www.nbcnews.com/tech/misinformation/taylor-swift-nude-deepfake-goes-viral-x-platform-rules-rcna135669].

64. *See supra* Part I.C.

B. LOW BARRIERS TO ENTRY FOR VIOLATIVE PLATFORMS

The availability of free open-source models and detailed instructional guides on web forums means nudification websites can be easily created and recreated.[65] The substantial revenue generated—driven by millions of monthly visitors—incentivizes continued operation despite legal risks.[66] It is easy to see why: a recent 2025 analysis of eighty-five such websites found that they averaged 18.5 million visitors monthly over a six-month period, with revenue estimates reaching as high as $36 million per website annually.[67] Nor is it especially easy to uncover the identities of the owners and operators: Registry information for these websites is frequently falsified or outdated, and many operate from overseas jurisdictions, complicating enforcement when content affects U.S. residents but originates from servers in Croatia, Estonia, China, or other countries.[68]

IV. LEGAL ENFORCEMENT OBSTACLES

In addition to the practical challenges of reining in deepfake NCII, certain legal obstacles—namely, section 230 and First Amendment protection—may also, at the margins, discourage enforcement. But while the invocation of section 230 and First Amendment defenses frequently looms large over efforts at legal regulation of many forms of deepfakes, such concerns are more remote in the context of NCII content.

A. SECTION 230 IMMUNITY CONCERNS

As noted above, section 230 has long played a central role in limiting enforcement of prohibitions on content creation and distribution online, and while the Take It Down Act has at least partially shifted this landscape, the precise contours remain to be determined. While section 230 immunity may not shield websites explicitly advertising nudification services—as such platforms clearly function as content contributors rather than neutral hosts—ambiguity persists regarding distribution platforms. The Take It Down Act attempts to address these concerns by creating specific takedown obligations that platforms must meet, with FTC enforcement authority treating violations as

65. *See* Hawkins et al., *supra* note 9.

66. *See* Lakatos, *supra* note 12.

67. Matt Burgess, *AI "Nudify" Websites Are Raking in Millions of Dollars*, WIRED (July 14, 2025), https://www.wired.com/story/ai-nudify-websites-are-raking-in-millions-of-dollars/ [https://web.archive.org/web/20260211200950/https://www.wired.com/story/ai-nudify-websites-are-raking-in-millions-of-dollars/].

68. Kolina Koltai, *Behind a Secretive Global Network of Non-Consensual Deepfake Pornography*, BELLINGCAT (Feb. 23, 2024), https://www.bellingcat.com/news/2024/02/23/behind-a-secretive-global-network-of-non-consensual-deepfake-pornography/ [https://web.archive.org/web/20260211201407/https://www.bellingcat.com/news/2024/02/23/behind-a-secretive-global-network-of-non-consensual-deepfake-pornography/] (finding that deepfake NCII websites are "incorporated in ways to hide the identity of their operators" and use virtual office services, with multiple companies using fake business addresses and falsely claiming partnerships with legitimate companies like Microsoft and G2A).

unfair or deceptive practices.[69] As noted, questions remain about whether and how section 230 protections interact with the Act's enforcement mechanisms.[70]

Uncertainty also persists regarding whether state-law identity rights claims—including right of publicity and NCII claims—fall within section 230's exception for intellectual property laws. Federal courts have reached conflicting conclusions on this question, with some holding that section 230 bars state publicity claims while others permit them to proceed.[71] This uncertainty complicates platform liability strategies and creates inconsistent protection for victims depending on jurisdiction.

B. FIRST AMENDMENT CONSIDERATIONS

Although First Amendment concerns frequently loom large in the context of regulating expressive conduct, free-speech challenges appear less formidable in this context than for other kinds of deepfake content. Obscenity and defamation have long constituted recognized exceptions to First Amendment protection,[72] and CSAM receives no constitutional protection at all. Obscenity is unprotected because it lacks serious literary, artistic, political, or scientific value and appeals to prurient interests in a patently offensive manner.[73] Similarly, defamation, which involves false statements that harm an individual's reputation, is generally excluded from First Amendment protection where the individual is a private figure (or, in the case of a public figure, where actual malice is proven).[74] CSAM, on the other hand, is categorically unprotected due to its intrinsic connection to the exploitation and abuse of children.[75] These exceptions reflect the U.S. Supreme Court's longstanding principle that certain types of speech, due to their harmful nature and lack of societal value, do not warrant constitutional safeguards.

The Supreme Court has yet to decide whether NCII—whether real or deepfake—is obscene, and therefore outside the protections of the First Amendment. But challenges to so-called "revenge porn" statutes have failed even where courts have declined to find such content categorically obscene. For example, in *State v. VanBuren*, the Vermont Supreme Court declined to label all nonconsensual pornography "obscene," but nevertheless, applying strict scrutiny, upheld a Vermont statute criminalizing its

69. Take It Down Act §§ 3(a)(3), 3(b).

70. *See* James Grimmelmann, Deconstructing *the Take It Down Act*, COMMC'NS OF THE ACM (July 30, 2025), https://cacm.acm.org/opinion/deconstructing-the-take-it-down-act/ [https://web.archive.org/web/20260211201709/https://cacm.acm.org/opinion/deconstructing-the-take-it-down-act/].

71. *Compare* Hepp v. Facebook, 14 F.4th 204 (3d Cir. 2021) (holding that Section 230 does not bar state right of publicity claim) *with* Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102 (9th Cir. 2007) (holding that Section 230 immunity barred right of publicity claim).

72. *See* Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 791 (2011).

73. *See* Miller v. California, 413 U.S. 15, 39 (1973).

74. *See, e.g.*, New York Times Co. v. Sullivan, 376 U.S. 254, 279–280 (1964); Gertz v. Robert Welch, Inc., 418 U.S. 323, 345–50 (1974).

75. *See* New York v. Ferber, 458 U.S. 747, 759–60 (1982); United States v. Williams, 553 U.S. 285, 288 (2008).

dissemination as narrowly tailored to serve a compelling State interest.[76] By contrast, the Illinois Supreme Court went even further, upholding Illinois's law under *intermediate scrutiny*, on the grounds that the prohibition was content-neutral and the statute regulated purely private matters and not speech on matters of public concern.[77]

More problematic are statutory prohibitions that sweep up protected content alongside that which falls outside the First Amendment's safeguards, because statutes that prohibit a substantial amount of protected expression may be deemed unconstitutionally "overbroad," failing even facial review.[78] Because most deepfake NCII statutory prohibitions have been enacted just within the past several years, the precise First Amendment boundaries remain unsettled and will be defined over time. Nevertheless, it is likely most will survive for the same reason nonconsensual pornography bans have: The state has a compelling interest in prohibiting the content and possesses few other available means to do so. The more interesting question is likely potential edge cases swept up by such bans: say, involving an artistic depiction of adults or newsworthy content—such as a journalist photographing a topless protester—that may technically constitute NCII but serve legitimate societal purposes or concern matters of public interest. Here, the risk would come not from whether banning NCII content *itself* is impermissibly overbroad, but whether if doing so inadvertently sweeps up enough protected speech alongside it that could succeed on an as-applied basis.

C. JURISDICTIONAL LIMITATIONS

Another significant enforcement challenge comes from the international scope of deepfake NCII content generation. Cross-border enforcement remains persistently problematic when websites or applications operate from foreign jurisdictions but generate content for users in the United States. The cross-border problem is not just that deepfake NCII content is unlawful in some places but under-regulated in others. Rather, the full abuse chain is typically transnational: The model or service has been developed and operated in one country, hosted in another, prompted by a user in a third, distributed via globally available platforms, and viewed wherever the victim may live. That fragmentation creates several recurring legal and practical obstacles.

V. STRATEGIC APPROACHES FOR ENHANCED ENFORCEMENT

Between the Take It Down Act and the kaleidoscope of recently enacted state laws, avenues for potential enforcement are theoretically vast. Notwithstanding the abundance of laws, however, there remains a deficit of enforcement.

A. PRIORITIZING GOVERNMENT AND PLATFORM ENFORCEMENT OVER INDIVIDUAL

76. State v. VanBuren, 214 A.3d 791, 800 (2019), as supplemented (June 7, 2019).
77. People v. Austin, 155 N.E.3d 439, 458 (2019).
78. Ashcroft v. Free Speech Coal., 535 U.S. 234, 244 (2002).

ACTIONS

At present, individual victim enforcement faces significant obstacles, even in seemingly straightforward cases. The experience of "Jane Doe," a New Jersey teenager, illustrates these challenges vividly. In October 2023, Doe discovered that a classmate, "K.G.," had used an AI nudification website, Clothoff.io, to generate deepfake nude images from her fully clothed Instagram photos and distributed them via Snapchat.[79] Though Doe successfully identified her perpetrator—an exceptional circumstance given that most victims never discover who created their images—enforcement barriers persisted. Criminal charges were not pursued because information gathered by school officials could not be used in the criminal investigation, and the defendant and witnesses refused to cooperate with law enforcement or provide access to their electronic devices.[80]

In February 2024, Doe filed a federal lawsuit against K.G. seeking $150,000 per image, injunctive relief, and destruction of all copies.[81] The psychological toll for Doe was devastating: Doe experienced "enormous distress" that "disrupted her high school education," living with "hopelessness and perpetual fear" knowing the images will "almost inevitably make their way onto the Internet."[82] Even this partial success—obtaining civil recourse against an identified perpetrator in close proximity—left the underlying platform problem unaddressed.

Thus, in October 2025, Doe filed a second lawsuit, this time targeting Clothoff.io itself, which Doe alleged is operated by AI/Robotics Venture Strategy 3 Ltd. in the British Virgin Islands.[83] Doe's second complaint includes both state and federal CSAM claims as well as state invasion of privacy claims, clearly illustrating the overlapping patchwork quilt of legal prohibitions against creating and disseminating NCII deepfake content. However, Doe's second complaint also illustrates the difference between legal rules on the books and effective enforcement on the ground: The complaint reveals the profound difficulties of pursuing overseas operators who use "pseudonyms, fake names and addresses, and third-party payment options to avoid detection and legal accountability," allegedly operating through Belarus-based individuals.[84] The procedural history of Doe's multi-year legal battle to seek justice demonstrates that even when victims overcome initial identification hurdles, cross-border jurisdiction and platform anonymity render individual enforcement anything but straightforward.

In this author's view, these challenges counsel toward systemic government enforcement mechanisms to support and complement individual victim lawsuits, an

79. Complaint at 4–5, Doe v. K.G., No. 2:24-cv-00634 (D.N.J. Feb. 2, 2024).

80. *Id.* at 11–12.

81. *Id.* at 28–29.

82. *Clinics File Suit Against Website that Generates Nonconsensual Nude Images*, YALE L. SCH. (Nov. 4, 2025), https://law.yale.edu/yls-today/news/clinics-file-suit-against-website-generates-nonconsensual-nude-images [https://web.archive.org/web/20260211202247/https://law.yale.edu/yls-today/news/clinics-file-suit-against-website-generates-nonconsensual-nude-images]; *id.* at 13.

83. Complaint at 5, Doe v. AI/Robotics Venture Strategy 3 Ltd., No. 2:25-cv-16671 (D.N.J. Oct. 16, 2025).

84. *Id.* at 30–42; *Clinics File Suit*, *supra* note 82.

approach recently-enacted laws like the Take It Down Act and California's amended NCII laws anticipate. One notable example of this approach is the August 2024 lawsuit filed by San Francisco City Attorney David Chiu against sixteen nudification websites, which collectively received over 200 million visits in the first half of 2024.[85] The suit, brought on behalf of the People of the State of California under the state's unfair competition law, seeks to shut down these websites and permanently enjoin their operation based on violations of state and federal revenge pornography, child pornography, and deepfake laws.[86] One targeted site, Clothoff.io—the same platform used against Jane Doe—had received approximately 26.9 million visits alone.[87] Since the filing of the People's suit, a majority of the sites are no longer accessible,[88] and two have entered into settlements with the People agreeing to permanently shut down the websites.[89]

As noted, the pending enforceability of the Take It Down Act's takedown provisions leaves open the broader question of platform liability—targeting not just content-creation platforms like nudification websites, but also platforms like Facebook and Snapchat. While such an approach has the potential to do much more to address the broader problem of systemic generation and viral dissemination, rather than isolated incidents with individual victims, much remains depending on how the FTC chooses to enforce the Act, how platforms respond, and how courts assess potential availability of section 230 defenses.

B. ENHANCING MODEL DISTRIBUTION REGULATIONS

While potentially controversial among open-source advocates, the emergence of NCII deepfakes also presents a cautionary tale for model distribution practices and raises important questions about the dangers of releasing open-source gen AI models. A proverbial Pandora's box was opened when Stable Diffusion 1.5 provided a sufficiently sophisticated open-source model without adequate retraining guardrails.[90] Since November 2022, the Civitai platform alone has hosted over 34,000 downloadable deepfake model variants, which collectively have been downloaded more than 15

85. First Amended Complaint, *supra* note 11, at 2, 4–6.

86. *Id.* at 20–24; *see* Heather Knight, *San Francisco Moves to Lead Fight Against Deepfake Nudes*, N.Y. TIMES (Aug. 15, 2024), https://www.nytimes.com/2024/08/15/us/deepfake-pornography-lawsuit-san-francisco.html.

87. Complaint at ¶ 136, People v. Sol Ecom, Inc., No. CGC-24-617237, WL 3833798 (Cal. Super. Ct. filed Aug. 14, 2024).

88. Video posted by David Chiu, FACEBOOK, *It has been one year since my Office filed our first-of-its-kind lawsuit against websites that use AI to create deepfake pornography.* (Aug. 28, 2025), https://www.facebook.com/davidchiu.sf/videos/it-has-been-one-year-since-my-office-filed-our-first-of-its-kind-lawsuit-against/745058168296841/ [https://web.archive.org/web/20260207004341/https://www.facebook.com/davidchiu.sf/videos/it-has-been-one-year-since-my-office-filed-our-first-of-its-kind-lawsuit-against/745058168296841/].

89. *See* Stipulated Judgments of May 30, 2025 and Dec. 19, 2025, People v. Sol Ecom, Inc., No. CGC-24-617237 (Cal. Super. Ct. filed Mar. 10, 2025).

90. Harris & Willner, *supra* note 8.

million times.[91] Video technology will rapidly and inevitably reach similar sophistication, and without protective measures, the current image-based crisis will extend to video content.

The challenge also extends beyond NCII to broader questions about identity rights and commercialization. Open-source models that enable unauthorized replication of voices and likenesses threaten not only privacy and dignity interests, but also the commercial value of identity for performers, athletes, and public figures. Policies addressing these concerns must carefully balance innovation benefits against potential harms, recognizing that overly restrictive approaches may stifle legitimate creative and educational uses, while overly permissive rules will fail to prevent determined bad actors from accessing or creating harmful tools.

Policymakers should consider several approaches to reducing the risk of open-source generative AI models contributing to NCII deepfakes. First, all platforms should commit to mandatory safety testing and red-teaming[92] before model release to minimize the risk of unanticipated usage or consequences and to ensure model guardrails cannot be broken.[93] Second, models should include built-in technical safeguards against malicious fine-tuning (although particularly with open-source models, current techniques have proven circumventable, and questions remain about the durability of open-source safeguards).[94] Especially critical are provisions allowing

91. Hawkins, *supra* note 9, at 1603.

92. Red-teaming is a process for evaluating and testing gen AI models to discover vulnerabilities, flaws, and unexpected outputs. To do so, the red-teamers often attempt to generate precisely the kind of content the model was designed to prohibit, either through manual prompt manipulation or use of specialized software that can engage with that model directly." Evelyn Yee, AI Red-Teaming Design: Threat Models and Tools, Oct. 24, 2025, https://cset.georgetown.edu/article/ai-red-teaming-design-threat-models-and-tools/ [http://web.archive.org/web/20260226184716/https://cset.georgetown.edu/article/ai-red-teaming-design-threat-models-and-tools/].

93. *See* Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Oct. 30, 2023) (mandating red-teaming for high-risk AI systems); Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [hereinafter the "EU AI Act"], arts. 15, 17, 2024 O.J. (L 1689) 1, 61 (requiring providers of high-risk AI systems to conduct testing and validation before deployment); *see also The Future of AI Red Teaming: Challenges, Trends, and What's Next*, AYA DATA (Nov. 13, 2025), https://www.ayadata.ai/the-future-of-ai-red-teaming-challenges-trends-and-whats-next/ [https://web.archive.org/web/20260207015645/https://www.ayadata.ai/the-future-of-ai-red-teaming-challenges-trends-and-whats-next/] (explaining that EU AI Act requires operators of high-risk AI systems to demonstrate accuracy and robustness through rigorous testing, and U.S. Executive Order mandates red teaming with safety test results shared with government agencies before deployment).

94. *See Second Key Update: Technical Safeguards and Risk Management*, INT'L AI SAFETY REP. (Nov. 25, 2025), https://internationalaisafetyreport.org/publication/second-key-update-technical-safeguards-and-risk-management/ [https://web.archive.org/web/20260207022050/https://internationalaisafetyreport.org/publication/second-key-update-technical-safeguards-and-risk-management] (noting that research explores "unlearning" techniques and methods to modify how models process harmful concepts, but such modifications "can often be reversed by actors with the technical skill to fine-tune models"); Xiangyu Qi et al., *On Evaluating the Durability of Safeguards for Open-Weight LLMs,* INT'L CONF. ON LEARNING REPRESENTATIONS 2025 1, 1 (2025) https://proceedings.iclr.cc/paper_files/paper/2025/file/9d3a4cdf6f70559e8c6fe02170fba568-Paper-Conference.pdf [https://perma.cc/QR48-6TWP] (demonstrating through case studies that evaluating

platforms to disable models found to facilitate widespread abuse, potentially backed by liability safe harbors for good-faith moderation efforts.[95] This is especially critical because once open-source models without remote disabling technology become available, it becomes nearly impossible to put the cat back in the bag.[96] Though current open-source models like Stable Diffusion primarily generate static images and lack the video generation capabilities of proprietary systems, more advanced open-source video platforms are quickly emerging, and will pose even greater risks unless preventive frameworks are established. These measures should be incorporated into risk-based regulatory frameworks that distinguish between different capability levels and use cases, as contemplated in the EU AI Act. Such an approach, for example, might distinguish between models than can generate only text-based outputs and those that can also generate images, video, or audio.[97]

C. CROSS-BORDER COOPERATION

Given practical and legal hurdles to worldwide enforcement and persistent extraterritoriality issues, enhanced international cooperation remains essential for meaningful progress. Comparative international approaches reveal both promising approaches and persistent gaps. The United Kingdom's Online Safety Act 2023, implemented over the following two years, criminalizes both the sharing and creation

technical safeguards is "exceedingly difficult" and defenses may mislead audiences about durability); Alex Petropoulos, Bengüsu Özcan & Max Reddel, *Can Open-Weight Models Ever Be Safe?*, CTR. FOR FUTURE GENERATIONS (Sep. 18, 2025), https://cfg.eu/can-open-weight-models-ever-be-safe/ [https://web.archive.org/web/20260207023543/https://cfg.eu/can-open-weight-models-ever-be-safe/] (discussing tamper-resistant architectures and machine unlearning techniques, but noting researchers have criticized these approaches for "either weakening useful capabilities or lacking resilience to circumvention"); Wiggers, *supra* note 10 ("Safety Classifier—while on by default—can be disabled.").

95. *See, e.g.*, Robert Gorwa & Michael Veale, *Moderating Model Marketplaces: Platform Governance Puzzles for AI Intermediaries*, 16 L., INNOVATION & TECH. 341, 341 (2024) (examining how Hugging Face, GitHub, and Civitai moderate models, including use of licensing, access restrictions, and automated content moderation).

96. *See* Martin Anderson, *CivitAI Tightens Deepfake Rules Under Pressure From Mastercard and Visa*, UNITE.AI (May 20, 2025), https://www.unite.ai/civitai-tightens-deepfake-rules-under-pressure-from-mastercard-and-visa/ [https://perma.cc/TT6M-AYD2] (reporting that Civitai banned models designed to replicate real people and added mandatory 50% noise alteration for uploaded images after payment processor pressure); *but see* Emanuel Maiberg, *Hugging Face Is Hosting 5,000 Nonconsensual AI Models of Real People*, 404 MEDIA (July 15, 2025), https://www.404media.co/hugging-face-is-hosting-5-000-nonconsensual-ai-models-of-real-people/ [https://web.archive.org/web/20260207030425/https://www.404media.co/hugging-face-is-hosting-5-000-nonconsensual-ai-models-of-real-people/] (documenting that over 5,000 models designed to create nonconsensual sexual content were reuploaded to Hugging Face after Civitai banned them).

97. *See* EU AI Act, *supra* note 93, arts. 5–6 (establishing prohibited AI practices and classification rules for high-risk AI systems based on their purpose and potential impact); *see also* Haiman Wong, *Mapping the Open-Source AI Debate: Cybersecurity Implications and Policy Priorities*, R ST. INST. (Apr. 17, 2025), https://www.rstreet.org/?post_type=research&p=85817 [https://perma.cc/JMU9-UT2Q] (discussing hybrid approaches that balance transparency with rigorous oversight, noting Meta's Llama model "requires users to apply for access and enforces a license that explicitly prohibits high-risk applications"); Petropoulos, *supra* note 94 (arguing for "tiered approach to evaluating the risks of an open-weight model at a given capability level and deciding whether it can be released safely under current conditions").

of sexually explicit deepfakes without consent, with penalties including unlimited fines and up to two years imprisonment.[98] The Act designates intimate image abuse as "priority illegal content," requiring covered platforms to implement systems for removal and empowering Ofcom—the UK communications regulator—to seek court orders compelling internet service providers to withdraw services from non-compliant sites.[99] Similarly, France amended its Code pénal in 2024 to criminalize non-consensual sexual deepfakes, imposing penalties of up to two years imprisonment and €60,000 fines.[100]

The European Union's AI Act, which entered into force in 2024, mandates transparency for AI-generated content and outlaws the most egregious forms of AI-based identity manipulation.[101] And the General Data Protection Regulation (GDPR)[102] continues to provide protection for depicted individuals' personal data: For instance, a Dutch court, citing the GDPR, recently ordered X/Grok to stop producing AI-generated non-consensual sexualized imagery, which extends even beyond a ban on NCII to include "sexualized" but non-nude depictions.[103] However, the EU approach focuses primarily on disclosure requirements and platform obligations rather than criminal penalties for individual creators. The European Commission's proposed regulation on child sexual abuse explicitly addresses deepfake CSAM, though implementation remains pending.[104]

These varied approaches highlight both the international consensus that NCII deepfakes require intervention and the lack of harmonized enforcement strategies. Effective cross-border cooperation requires more than parallel national laws—it

98. *Criminalising Deepfakes—The UK's New Offences Following the Online Safety Act*, HERBERT SMITH FREEHILLS KRAMER (May 21, 2024), https://www.herbertsmithfreehills.com/notes/tmt/2024-05/criminalising-deepfakes-the-uks-new-offences-following-the-online-safety-act [http://web.archive.org/web/20240619083113/https://www.herbertsmithfreehills.com/notes/tmt/2024-05/criminalising-deepfakes-the-uks-new-offences-following-the-online-safety-act]; Press Release, Sarah Sackman, MP, Ministry of Just. (UK), Better Protection for Victims Thanks to New Law on Sexually Explicit Deepfakes (Jan. 22, 2025), https://www.gov.uk/government/news/better-protection-for-victims-thanks-to-new-law-on-sexually-explicit-deepfakes [https://web.archive.org/web/20260207040234/https://www.gov.uk/government/news/better-protection-for-victims-thanks-to-new-law-on-sexually-explicit-deepfakes].

99. WOMEN AND EQUALITIES COMMITTEE, TACKLING NON-CONSENSUAL INTIMATE IMAGE ABUSE: GOVERNMENT RESPONSE, 2024-5, HC 911, at 4, 20 (UK).

100. Henry Patishman, *Global Legal Actions Against AI Deepfakes: Five Laws of 2025*, REGULA (Aug. 12, 2025), https://regulaforensics.com/blog/deepfake-regulations/ [https://web.archive.org/web/20260207041909/https://regulaforensics.com/blog/deepfake-regulations/].

101. EU AI Act, *supra* note 93, art. 50, (requiring transparency obligations for AI-generated content).

102. Commission Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

103. *See* Ramsha Jahangir, *Dutch Court Orders X, Grok to Stop AI-Generated Sexual Abuse Content*, TECH POLICY PRESS (Mar. 26, 2026) https://www.techpolicy.press/dutch-court-orders-x-grok-to-stop-aigenerated-sexual-abuse-content/ [https://web.archive.org/web/20260330131413/https://www.techpolicy.press/dutch-court-orders-x-grok-to-stop-aigenerated-sexual-abuse-content/].

104. HERBERT SMITH FREEHILLS KRAMER, *supra* note 98.

demands mutual legal assistance, treaties adapted to digital evidence, coordinated platform takedown procedures, and mechanisms for pursuing operators who deliberately forum-shop across jurisdictions. The *Jane Doe v. Clothoff* litigation exemplifies these challenges: An operator allegedly based in Belarus, incorporated in the British Virgin Islands, using fake addresses in Argentina, and inflicting harm on U.S. victims represents precisely the enforcement quagmire that demands international coordination.[105]

D. THE LIMITS OF TECHNICAL DETECTION SOLUTIONS

Although a technological solution would theoretically be the easiest remedy, technical approaches to identifying deepfakes have proven inadequate as standalone solutions. Digital watermarking—embedding invisible signatures into generated content—faces fundamental vulnerabilities. While platforms like Google's SynthID and Meta's StableSignature embed watermarks intended to survive common image manipulations, researchers have demonstrated that these protections can be systematically defeated. The "UnMarker" system developed by University of Waterloo researchers successfully removed watermarks from seven leading schemes, reducing detection accuracy below random-guess levels without access to training data or internal system details.[106]

Watermarking also suffers from inherent limitations beyond removal efforts. For watermarks to function, they must be universally adopted—yet open-source models and bad-faith actors presently face no obligation, legal or otherwise, to implement them. Watermarks embedded during generation cannot identify content created using models that predate watermarking requirements or that deliberately omit such features. Even when present, watermarks provide attribution rather than prevention; they identify synthetic content after dissemination—potentially mitigating the deceptive features of deepfake content—but do nothing to stop initial creation or distribution.[107]

Perceptual hashing and metadata embedding face similar constraints. While metadata can theoretically track content provenance, it remains easily stripped,[108]

105. Complaint, *supra* note 79, at 3, 5–6, 21.

106. Irfan Ahmad, *Researchers Break Industry Watermarks, Undermining Key Deepfake Detection Methods*, DIGIT. INFO. WORLD (July 24, 2025), https://www.digitalinformationworld.com/2025/07/researchers-break-industry-watermarks.html [https://web.archive.org/web/20260207045148/https://www.digitalinformationworld.com/2025/07/researchers-break-industry-watermarks.html].

107. *See* Nick Gaubitch, *Does Watermarking Protect Against Deepfake Attacks?*, PINDROP (Oct. 30, 2025), https://www.pindrop.com/article/does-watermarking-protect-against-deepfake-attacks/ [https://web.archive.org/web/20260207045515/https://www.pindrop.com/article/does-watermarking-protect-against-deepfake-attacks/].

108. *C2PA in ChatGPT Images*, OPENAI, https://help.openai.com/en/articles/8912793-c2pa-in-chatgpt-images [https://web.archive.org/web/20260228054757/https://help.openai.com/en/articles/8912793-c2pa-in-chatgpt-images] (last visited Feb. 21, 2026).

including by using the same AI tools that generated the content.[109] Perceptual hashing—creating digital "fingerprints" of images to identify duplicates—requires access to databases of known NCII images, operates reactively rather than preventively, and struggles with slight variations in manipulated content.[110] These technical measures may assist in enforcement by providing evidence of synthetic origins, but they cannot substitute for legal frameworks that impose liability on creators and platforms regardless of whether content bears detectable markers.

To date, the rapid evolution of generative AI has consistently outpaced technical countermeasures. Detection algorithms trained on current deepfake artifacts become obsolete as new generation techniques emerge. This arms race dynamic suggests that legal and economic interventions targeting creation and distribution infrastructure offer complementary—and potentially more durable—solutions than technical detection alone.

E. VOLUNTARY COOPERATION

Although important, neither government enforcement nor technical detection alone is likely to eliminate the threat of deepfake NCII content. Voluntary cooperation from key stakeholders also holds great promise as a part of a holistic approach to mitigating harm. Domain registrars, web hosts, and app stores each maintain their own terms of service, which almost always prohibit hosting CSAM content, frequently prohibit hosting NCII content, and often prohibit content and conduct that constitutes cyber-bulling, abuse, and harassment. While not traditional government enforcement mechanisms, private, quasi-contract-based remedies deserve exploration as well. Platforms have historically demonstrated acute sensitivity to CSAM liability concerns, making terms of service violations a potentially effective supplementary enforcement avenue for NCII content more broadly.

Economic pressure points offer additional leverage. Research reveals that just a subset of identified nudification websites collectively generated at least $36 million in annual revenue, supported by mainstream technology infrastructure.[111] Analysis of eighty-five such websites found that sixty-two utilize Amazon Web Services or Cloudflare for hosting and content delivery, while fifty-four employ Google's sign-on system.[112] These sites monetize through tiered subscription models, typically offering limited free trials before requiring payment—commonly accepting cryptocurrency to

109. Jacob Hoffman-Andrews, *AI Watermarking Won't Curb Disinformation*, ELEC. FRONTIER FOUND. (Jan. 5, 2024), https://www.eff.org/deeplinks/2024/01/ai-watermarking-wont-curb-disinformation [https://web.archive.org/web/20260207052221/https://www.eff.org/deeplinks/2024/01/ai-watermarking-wont-curb-disinformation].

110. *See* Hannes Mareen et al., *Fast Fallback Watermark Detection Using Perceptual Hashes*, 10 ELECS. 1155 (2021).

111. *See* Burgess, *supra* note 67.

112. *See id.*

evade financial oversight, but also processing payments through PayPal, Apple Pay, Cash App, Venmo, and traditional credit cards.[113]

This financial ecosystem supporting deepfake NCII content creation thus presents additional opportunities for intervention. Payment processors and financial services companies can refuse to process transactions for known nudification services, as they have done for other categories of harmful content.[114] Cloud infrastructure providers can terminate services to platforms violating their acceptable use policies.[115] While such private enforcement raises concerns about unchecked corporate power and potential overreach,[116] targeted actions against platforms whose core business model centers on facilitating illegal content production present compelling cases for intervention. Cryptocurrency's role in enabling these services—offering anonymity and circumventing traditional financial controls—highlights the need for regulatory attention to payment processing infrastructure supporting illegal content generation.

---

113. Complaint, *supra* note 79, at ¶¶s 41, 50, 59, 74, 82, 141 (detailing payment methods accepted by various nudification websites including PayPal, Apple Pay, Cash App, Venmo, cryptocurrency, and traditional credit cards).

114. *See e.g.*, Michelle Price, *Mastercard, Visa Suspend Ties with Ad Arm of Pornhub Owner MindGeek*, REUTERS (Aug. 4, 2022), https://www.reuters.com/business/finance/mastercard-visa-suspend-ties-with-ad-arm-pornhub-owner-mindgeek-2022-08-04/ [https://web.archive.org/web/20260206184939/https://www.reuters.com/business/finance/mastercard-visa-suspend-ties-with-ad-arm-pornhub-owner-mindgeek-2022-08-04/].

115. Major cloud infrastructure providers maintain acceptable use policies that authorize service termination for policy violations. *See AWS Acceptable Use Policy*, AMAZON WEB SERVS., https://aws.amazon.com/aup/ [https://web.archive.org/web/20260206191410/https://aws.amazon.com/aup/] (last visited Feb. 8, 2026) ("We may . . . remove or disable access to any content…that violates this Policy."); *Google Cloud Platform Terms of Service* § 4.1, GOOGLE CLOUD, https://cloud.google.com/terms [https://web.archive.org/web/20260206192644/https://cloud.google.com/terms] (last visited Feb. 8, 2026) ("Google may Suspend all or part of Customer's use of the Services until the violation is corrected."); *For Online Services*, MICROSOFT, https://www.microsoft.com/licensing/terms/product/ForOnlineServices/all [https://perma.cc/XRK9-T5G6] (last visited Feb. 8, 2026) ("[V]iolations of the Acceptable Use Policy . . . may result in suspension of the Online Service."); *Cloudflare Self-Serve Subscription Agreement*, CLOUDFLARE, https://www.cloudflare.com/terms/ [https://web.archive.org/web/20260206193559/https://www.cloudflare.com/terms/] (last visited Feb. 8, 2026) ("We may at our sole discretion terminate your user account or Suspend or terminate your use or access to the Service at any time, with or without notice for any reason or no reason at all."); *Acceptable Use Policy*, IDENTITY DIGIT., https://www.identity.digital/policies/acceptable-use-policy [http://web.archive.org/web/20260206193749/https://www.identity.digital/policies/acceptable-use-policy] (last visited Feb. 8, 2026) ("Identity Digital reserves the right . . . to deny, suspend, cancel, redirect, or transfer any registration or transaction . . . as it determines necessary . . . to comply with the terms of the applicable registration agreement and Identity Digital' policies [or where] . . . domain name use is abusive or violates this AUP.").

116. *See, e.g.*, *FIRE Statement on Free Speech and Online Payment Processors*, FOUND. FOR INDIVIDUAL RTS. & EXPRESSION (FIRE), https://www.thefire.org/research-learn/fire-statement-free-speech-and-online-payment-processors [http://web.archive.org/web/20260206194300/https://www.thefire.org/research-learn/fire-statement-free-speech-and-online-payment-processors] (last visited Feb. 8, 2026) ("When these companies appoint themselves the arbiters of what speech and views are acceptable, shutting people and organizations out of the online financial ecosystem for wrongthink, they seriously undermine our culture of free expression.").

VI. CONCLUSION

The recent explosion of AI-generated deepfake non-consensual intimate images demonstrates that even when a widespread societal consensus had concluded that a problem demands intervention, implementing enforcement to address it remain profoundly challenging. Practical obstacles abound: Technology makes the creation and distribution of NCII all too easy but the source attribution incredibly challenging. Cheap, readily replicable open-source models create a "whack-a-mole" problem at the platform level. These challenges, coupled with the jurisdictional limitations in enforcing an inherently cross-border phenomenon, make it likely that no single solution will prove sufficient.

Fortunately, a patchwork quilt of solutions stands at the ready, if political and technical know-how are properly employed. Effective responses require coordinated efforts across multiple domains: enhanced platform accountability, reconsideration of open-source model distribution practices, strengthened cross-border cooperation, and creative use of both public enforcement mechanisms and private contractual remedies. As generative AI capabilities continue advancing, the window for establishing protective frameworks before video-based NCID becomes as prevalent as current image-based content continues to narrow. Policymakers, technology companies, and civil society must act decisively while acknowledging that addressing this crisis will require sustained, adaptive efforts rather than singular legislative fixes.

ISSN 1544-4848
COLUM. J.L. & ARTS