Voice Cloning in an Age of Generative AI:
Mapping the Limits of the Law &
Principles for a New Social Contract with Technology

Paulius Jurcys, Mark Fenwick, and Aidas Liaudanskas

August 24, 2024

Abstract

The rise of generative AI technologies has enabled the development of sophisticated voice cloning capabilities, presenting significant ethical, social, and legal challenges. This paper explores these complexities and advocates for a new social contract and related legal principles that address the implications of such technologies. The authors examine the historical context of voice replication technologies, from early speech synthesis efforts to modern deep learning techniques like WaveNet and Tacotron 2, which have dramatically improved the realism and accessibility of voice cloning. Despite the technological advances, the current legal framework is deeply inadequate in recognizing and addressing the unique issues posed by voice AI. The paper critiques existing laws related to copyright, privacy, and publicity rights, highlighting their limitations in protecting individuals from unauthorized voice cloning. The authors argue that the fundamental principles of digital self-determination and individual consent must be established as the cornerstones of new regulations. The paper proposes practical recommendations, including implementing default consent principles, transparency in AI-generated content, and developing industry standards and federal legislation such as the proposed Federal Anti-Impersonation Right Act. Such principles can provide a more robust basis to ensure that individuals maintain control over their voices, address widespread concerns around voice cloning technologies, and foster a balanced relationship between technological innovation and individual rights, security, and well-being.

Keywords

Table of Contents

1. Introduction

On May 13, 2024, OpenAI released its ChatGPT-4o model, which included a new voice communication interface. The new model featured five different voice tones, one of which—a female-sounding voice named "Sky"—was quickly noted by commentators to resemble Hollywood actress Scarlett Johansson's voice as found in her AI portrayal in the 2013 film "Her."[1]

Soon after the release, Ms. Johansson published a statement on X (formerly Twitter), accusing OpenAI and its founder, Sam Altman, of deliberately copying her voice.[2] She expressed her disappointment and frustration at the similarity between the voice-powered functionality of ChatGPT-4o and her voice: "When I heard the released demo, I was shocked, angered, and in disbelief that Mr. Altman would pursue a voice that sounded so eerily similar to mine."[3] Johansson also indicated that even her closest friends could not tell the difference between her voice and "Sky," one of the voice iterations of ChatGPT-4o.[4]

In her statement, Johansson explained that OpenAI CEO Sam Altman initially approached her in September 2023 about becoming the voice for "Sky," but she declined. In an open letter, Johansson noted that "[Sam Altman] told me that he felt that by my voicing the system, I could bridge the gap between tech companies and creatives and help consumers feel comfortable with the seismic shift concerning humans and AI."[5] Months later, two days before the release of ChatGPT-4o, Sam Altman contacted her agent again, urging Johansson to reconsider her initial refusal to cooperate.[6] "Mr. Altman even insinuated the similarity was intentional, tweeting[7] a single word "her"—a reference to the film in which I voiced a chat system, Samantha, who forms an intimate relationship with a human."

Later, Altman apologized on X, stating that OpenAI could provide evidence that they didn't use her voice directly: "The voice of Sky is not Scarlett Johansson's, and it was never intended to resemble hers. We cast the voice actor behind Sky's voice before any outreach to Ms. Johansson. Out of respect for Ms. Johansson, we have paused using Sky's voice in our products. We are

[1] Bobby Allyn, *Voice Analysis Shows Striking Similarity Between Scarlett Johansson and ChatGPT*, NPR, https://www.npr.org/2024/05/31/g-s1-2263/voice-lab-analysis-striking-similarity-scarlett-johansson-chatgpt-sky-opennai(last visited Aug. 22, 2024).

[2] Bobby Allyn, *Voice Analysis Shows Striking Similarity Between Scarlett Johansson and ChatGPT*, NPR, https://www.npr.org/2024/05/31/g-s1-2263/voice-lab-analysis-striking-similarity-scarlett-johansson-chatgpt-sky-opennai(last visited Aug. 22, 2024).

[3] *BestfScarlett*, Twitter (Mar. 12, 2024, 3:45 PM), https://twitter.com/BestfScarlett/status/1792688497272512739.

[4] *BestfScarlett*, Twitter (Mar. 12, 2024, 3:45 PM), https://twitter.com/BestfScarlett/status/1792688497272512739.

[5] *BestfScarlett*, Twitter (Mar. 12, 2024, 3:45 PM), https://twitter.com/BestfScarlett/status/1792688497272512739.

[6] *BestfScarlett*, Twitter (Mar. 12, 2024, 3:45 PM), https://twitter.com/BestfScarlett/status/1792688497272512739.

[7] *BestfScarlett*, Twitter (Mar. 12, 2024, 3:45 PM), https://twitter.com/BestfScarlett/status/1792688497272512739.

sorry to Ms. Johansson that we didn't communicate better."[8] OpenAI issued a separate statement explaining that the voice wasn't an intentional imitation and that it came from an actor they hired who was speaking in their normal voice.[9] However, OpenAI refused to share the actor's name, citing privacy concerns.

This incident made headlines in the media around the world and captured the attention of many public commentators, including leading VCs in Silicon Valley.[10] OpenAI's controversy was even featured on Saturday Night Live.[11] Most of the public conversation around the incident appeared to be sympathetic to Scarlett Johansson, and the suggestion was that OpenAI had behaved badly, or at least disingenuously.

Yet, as generative AI technologies proliferate, disputes over voice cloning and AI-powered music generation are likely to increase.[12] Days after Scarlet Johannson's dispute with OpenAI, another VC-funded synthetic speech-generation company, LOVO, made headlines because of the unauthorized use of people's voices. 13 The firm was hit with a class action alleging the company misappropriated voiceover actors' voices and deceptively marketed its product.[13]

Using copyrighted text data from printed publications to train language-based generative AI models has already been scrutinized by national courts and international policy-making bodies.[14] In contrast, the use of scraped data to replicate voices and develop voice cloning technologies has not yet garnered as much attention.[15] Arguably, the stakes are much higher when it comes to the unauthorized use of a person's voice in various generative AI scenarios. This is because there

---

[8] OpenAI, *How the Voices for ChatGPT Were Chosen*,
https://openai.com/index/how-the-voices-for-chatgpt-were-chosen/(last visited Aug. 22, 2024).

[9] OpenAI, *How the Voices for ChatGPT Were Chosen*,
https://openai.com/index/how-the-voices-for-chatgpt-were-chosen/(last visited Aug. 22, 2024).

[10] E.g., All In Podcast, *Scarlett Johansson vs. OpenAI, Nvidia's Trillion-Dollar Problem, a Vibecession, Plastic in Our Balls* (May 24, 2024), https://www.youtube.com/watch?v=c9HEjdjyVn4.

[11] See Saturday Night Live, *Weekend Update: Colin Jost and Michael Che Swap Jokes for Season 49 Finale - SNL*, https://www.youtube.com/watch?v=HPH0HgotIE4.

[12] See e.g., openAI, *We're Starting to Roll Out Advanced Voice*, LinkedIn, https://www.linkedin.com/posts/openai_were-starting-to-roll-out-advanced-voice-activity-7224119544313733121-V hW1?utm_source=share&utm_medium=member_desktop (last visited Aug. 22, 2024); Suno, *An AI Music Generator "Where Anyone Can Make Music,"* Suno, https://suno.com/?utm_source=theneuron (last visited Aug. 22, 2024).

[13] See *Voice Cloning Class Action*, Pollock Cohen LLP,
https://www.pollockcohen.com/cases-investigations/was-your-voice-cloned-and-sold-by-lovo (last visited Aug. 22, 2024).

[14] See, e.g., Paulius Jurcys & Mark D. Fenwick, *NY Times vs. Microsoft and OpenAI: Should It Be an "Easy" Fair Use Case to Decide?*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4685275 (last visited Aug. 22, 2024).

[15] See, e.g., Daniel J. Solove & Woodrow Hartzog, *The Great Scrape: The Clash Between Scraping and Privacy*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4884485 (last visited Aug. 22, 2024).

is closer proximity between an individual's voice and their identity, and one could reasonably argue that our voices are more intimately connected to our identities than any text-based documents we create.

Not surprisingly, AI-powered voice cloning technologies raise numerous legal, ethical, social, and economic questions. One might question whether OpenAI behaved ethically by releasing "Sky"—knowing that it would cause confusion and that it sounded eerily similar to Scarlett Johansson's voice. While not illegal, it is clearly unethical or, at the very least, ethically dubious. From a legal perspective, the first sanction in such a scenario is the social shaming and stigma that OpenAI now faces, making the company appear clumsy and indicative of a poor corporate culture.[16] The voice AI industry has reacted with disappointment, as OpenAI's selfish behavior portrays a damaging picture of other industry players who are trying to act more cautiously and make a positive impact with voice AI.

While the idea of perfectly cloning someone's voice may sound gimmicky or simply unnerving, synthetic voice technologies have a wide range of practical and socially beneficial commercial applications.[17] For example, voice clones allow text-to-speech interfaces to use personalized, human-sounding voices rather than generic synthesized ones. Voice AI can also facilitate human inclusion. For example, people who have lost their voices due to illness or injury might use a cloned version of their own voice. Similarly, voice AI technologies could be used for audiobook narration, podcasting, video game voice acting, and more. In business-to-consumer services, voice AI could be useful for call centers or other customer-facing services. Using voice AI technologies, it is possible to match customer demographics for a more personalized and valuable experience.

Generative AI technologies are already having a profound impact on humanity. They can optimize processes, reduce transaction costs, increase access to knowledge and information, and open avenues for the creation of new business models and value streams. Yet, as voice cloning examples illustrate, society has not been adequately prepared for the arrival of generative AI.[18] This is particularly evident in the realm of public policy and regulation, where existing legal concepts, principles, and doctrines do not always provide a sufficiently clear and robust

---

[16] See (1) the quotes from the All In Podcast, *Scarlett Johansson vs. OpenAI, Nvidia's Trillion-Dollar Problem, a Vibecession, Plastic in Our Balls*(May 24, 2024), https://www.youtube.com/watch?v=c9HEjdjyVn4., supra note. OpenAI keeps making those enormous blunders. (2) Another example is with the NY Times, while we are sympathetic to liberal values of the NY Times, OpenAI seems to have blundered there as well.

[17] E. Meskys et al., *Regulating Deep Fakes: Legal and Ethical Considerations*, 15 J. Intell. Prop. L. & Prac. 24, 24–31 (2020).

[18] Paulius Jurcys et al., *Artificial Intelligence and Copyright: Comments on a Notice Requested by the U.S. Copyright Office*, https://www.regulations.gov/comment/COLC-2023-0006-9137 (last visited Aug. 22, 2024); Paulius Jurcys & Mark Fenwick, *Creativity and Information Intermediaries in the Age of Generative AI*, Forthcoming in *The International Journal of Information Management* (IJIM) 2024.

framework for addressing various AI-created legal issues and concerns. Accordingly, from a legal, social, and economic perspective, it is obvious that society needs to forge a new social contract with these emerging technologies, including robust regulatory principles.[19] While the possibilities of voice replication are not unprecedented, the rapid advancements in underlying technologies compel us to reassess and redefine the legal and ethical frameworks governing their use.

In Section 2, we detail the technological advancements that have rendered voice cloning remarkably accessible, noting that a mere fifteen to twenty seconds of an individual's voice is sufficient to create a convincing clone.[20] To initiate a meaningful dialogue on the necessity and form of a new social contract, we argue that it is imperative to first map existing legal principles. Section 3, therefore, examines various frameworks that govern creative expressions and protect privacy and publicity rights, reputation, consumers, and markets. We highlight the limitations of the current legal framework when dealing with issues of voice AI and suggest that individuals' voices have only ever been a tangential concern of regulatory interventions and that, consequently, the existing law fails to adequately address this issue.

These limitations in the extant legal imagination lead us to conclude that we need a new approach. Hence, our suggestion here is to begin thinking about a new social contract with technology and the fundamental principles that might then allow us to develop an appropriate regulatory framework and rules to address the unique challenges posed by voice AI.

In Section 4, we assert that a fundamental element of any such new social contract is respect for individual dignity. In a post-digital transformation world, an individual's voice needs to be acknowledged as a fundamental component of their identity. Our unique individual voice is integral to being recognized and distinguished in society, and that voice is an expressive characteristic that defines us as unique, free, and autonomous individuals. Accordingly, individuals should have full control over their voice, notably as an economic resource that can provide the basis of their economic livelihood, and such control should, therefore, be essential for any individual. Building on this idea, we advocate for the principle of digital self-determination, arguing that an individual's consent should, consequently, be considered as an expression of such control over their voice and how it is used in AI-powered environments. From a legal perspective, we contend that consent must be the foundational precondition for the deployment of voice AI technologies. This principle should apply universally, irrespective of whether the individual is a private citizen or a public figure. Individual consent should be entrenched as the golden rule. We observe that emerging industry standards, particularly in

---

[19] Paulius Jurcys & Mark Fenwick, *Creativity and Information Intermediaries in the Age of Generative AI*, Forthcoming in *The International Journal of Information Management* (IJIM) 2024.
[20] HyeGen and ElevenLabs websites.

direct-to-consumer products, are beginning to adopt this type of approach, allowing users to opt-in to clone their voices and accept the associated consequences.

Building on this premise, Section 5 proposes five high-level principles that could be useful in thinking about such new social contract with (voice AI) technologies, namely (i) respect for individual dignity, (ii) ownership, (iii) consent, (iv) transparency and disclosure, and (v) accountability and redress. This new social contract perspective highlights the limitations of current laws and enables us to explore the contours of new and necessary regulations.

2. The Technological Foundations of Voice Clones

By way of an introduction to the ethical and legal issues, it is helpful to begin with a brief introduction to the technological foundations of voice cloning, specifically the recent leaps forward that have occurred in this technology.

2.1. Early Attempts to Recreate Voice

The quest to replicate human speech synthetically began in the mid-20th century with early speech synthesis efforts.[21] These initial attempts focused on mimicking human speech patterns through rule-based synthesis during the period of 1930s-1980s. One notable project was the "Voder," introduced by Bell Labs at the 1939 World's Fair.[22] While it could produce recognizable speech sounds, it required manual operation, highlighting the complexity of speech synthesis. In the 1960s to 1980s, researchers developed more sophisticated text-to-speech systems. The formant synthesis-based "DECtalk," which gained popularity in the 1980s and was famously used by Stephen Hawking, represented a significant advancement in this field.[23] "DECtalk" used acoustic-phonetic rules to generate speech sounds.

---

[21] *What Is the History of Text to Speech and Voice Synthesis?*, Speechify, https://speechify.com/blog/history-of-text-to-speech-voice-synthesis/ (last visited Aug. 22, 2024).

[22] *What Is the Voder?*, https://www.whatisthevoder.com (last visited Aug. 22, 2024); Eric Grundhauser, *The Voder, the First Machine to Create Human Speech: It Spoke Like A Demon*, Atlas Obscura (Jan. 16, 2017), https://www.atlasobscura.com/articles/the-voder-the-first-machine-to-create-human-speech (last visited Aug. 22, 2024); Kat Eshner, *Meet Pedro the "Voder," the First Electronic Machine to Talk*, Smithsonian Mag., https://www.smithsonianmag.com/smart-news/meet-pedro-voder-first-electronic-machine-talk-180963516/#:~:text= Pedro%20was%20an%20experiment%20in%20reproducing%20speech%20electronically (last visited Aug. 22, 2024).

[23] Digital Equipment Corporation, *DECtalk DTC-01*, https://archive.org/details/dectalk#:~:text=DECtalk%20units%20could%20recognize%20and%20generate%20any% 20telephone%20touch%20tone(last visited Aug. 22, 2024).

The 1980s saw further progress with the development of the Klatt synthesizer at AT&T Bell Laboratories.[24] This system used a set of acoustic-phonetic rules to generate speech sounds, producing more natural-sounding output than its predecessors.[25]

In the 1990s, concatenative synthesis emerged as a new approach. This technique involved splicing together pre-recorded speech segments to form new utterances.[26] Concatenative synthesis relies on a comprehensive database of recorded speech segments, including phonemes, diphones, and triphones, to capture natural speech dynamics. When text input is received, the system analyzes the phonetic sequence and employs algorithms to select the best-matching speech units from the database. These units are then concatenated to form the final speech output, with smoothing techniques ensuring seamless transitions.

An example of concatenative synthesis in action is the AT&T Natural Voices text-to-speech (TTS) system, developed in the early 2000s.[27] This system was used in various applications, including telephony systems, assistive technologies, and multimedia content generation. The success of AT&T Natural Voices demonstrated the potential of concatenative synthesis to deliver high-quality TTS systems.

As speech synthesis technology evolved, researchers introduced a method called Hidden Markov Models (HMMs) in the late 1990s and early 2000s. This approach used statistical techniques to analyze and recreate the patterns of human speech. The HTS system, introduced by Tokuda et al.[28] in 2002, became a widely used framework for HMM-based speech synthesis. It demonstrated how statistical models could generate natural-sounding speech while offering greater flexibility than previous methods.

While these early attempts laid the groundwork for modern voice cloning, they had limitations. Concatenative synthesis, for instance, required extensive manual processing and editing.[29] Moreover, these techniques often failed to capture the nuances and natural variations of human speech, resulting in synthetic voices that sounded robotic and unnatural.[30] True voice cloning,

---

[24] Dennis H. Klatt, *Review of Text-to-Speech Conversion for English*, 82 J. Acoust. Soc. Am. 737, 737–793 (1987).

[25] ibid.

[26] Diemo Schwarz, *Concatenative Sound Synthesis: The Early Years*, 35 J. New Music Res. 3, 3–22 (2006).

[27] B.H. Juang & Lawrence R. Rabiner, *Automatic Speech Recognition – A Brief History of the Technology Development*, https://web.ece.ucsb.edu/Faculty/Rabiner/ece259/Reprints/354_LALI-ASRHistory-final-10-8.pdf (last visited Aug. 22, 2024).

[28] Keiichi Tokuda, Heiga Zen & Alan W. Black, *An HMM-Based Speech Synthesis System Applied to English*, in *IEEE Speech Synthesis Workshop* (Santa Monica, 2002).

[29] Diemo Schwarz, *Current Research in Concatenative Sound Synthesis*, in *Proceedings of the International Computer Music Conference* 1, 1 (Barcelona, Spain, Sept. 2005).

[30] Carol Moh, *How to Use Voice Cloning AI to Replicate Your Voice*, Lovo.ai (last visited Aug. 22, 2024), https://lovo.ai/post/how-to-use-voice-cloning-ai-to-replicate-your-voice.

which involves replicating individual voice characteristics, remained elusive with these early methods.

2.2. The Rise of Deep Learning and Its Impact on Voice AI

2010s witnessed the rise of Deep Learning. Beginning with simple Deep neural networks like the one by Zen et al.,[31] followed by recurrent networks e.g. by Fan et al.[32] Deep networks addressed some of the limitations of previous models by effectively learning the temporal dynamics of speech, including intonation and rhythm. This led to more natural-sounding synthesized speech with improved prosody.

WaveNet, introduced by Google in 2016,[33] was a major breakthrough in voice synthesis. It used a type of neural network called a convolutional neural network (CNN) to model raw audio waveforms directly. This allowed WaveNet to capture subtle details in speech, like intonation and pronunciation, resulting in much more natural-sounding synthetic voices. However, WaveNet had a significant drawback: it was computationally intensive and slow to generate audio.

The next big leap came in 2017 with the introduction of the Transformer architecture.[34] Originally designed for language translation tasks, Transformers quickly proved to be a game-changer across many areas of AI, including voice synthesis. Transformers were revolutionary due to their "attention" mechanism, which allows the model to focus on relevant parts of the input when generating output, leading to more accurate and natural-sounding speech. Unlike previous models that processed input sequentially, Transformers can handle entire sentences in parallel, making them much faster to train and use. They excel at understanding context over longer sequences, resulting in more coherent and contextually appropriate speech synthesis. Moreover, the Transformer architecture's versatility allowed it to be adapted for various tasks, including text-to-speech, voice cloning, and even controlling aspects like emotion and style in synthesized speech. These advantages made Transformers a cornerstone of modern

---

[31] H. Zen, A. Senior & M. Schuster, *Statistical Parametric Speech Synthesis Using Deep Neural Networks*, in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* 7962, 7962–7966 (Vancouver, Can., 2013), https://doi.org/10.1109/ICASSP.2013.6639215.

[32] Yuchen Fan et al., *TTS Synthesis with Bidirectional LSTM Based Recurrent Neural Networks*, in *Proceedings of Interspeech 2014* (2014).

[33] See *WaveNet: A Generative Model for Raw Audio*, DeepMind, https://deepmind.google/discover/blog/wavenet-a-generative-model-for-raw-audio/ (last visited Aug. 22, 2024) and *WaveNet*, DeepMind, https://deepmind.google/technologies/wavenet/ (last visited Aug. 22, 2024).

[34] Ashish Vaswani et al., *Attention Is All You Need*, in *Proceedings of the 31st International Conference on Neural Information Processing Systems* 6000, 6000–6010 (2017).

voice AI technologies, enabling significant improvements in the quality and capabilities of voice synthesis systems.

These advantages of Transformers quickly led to their adoption in voice AI. For example, Google's Tacotron 2[35] and NVIDIA's FastSpeech[36] leveraged Transformer-based architectures to create high-quality voice synthesis systems that could not only generate natural-sounding speech but also mimic specific voices with remarkable accuracy. This opened up new possibilities for applications ranging from more realistic virtual assistants to personalized audiobook narration. The impact of Transformers continued to grow, with subsequent models like Meta's VoiceBox (2023) combining Transformer architectures with other advanced techniques like diffusion models. These hybrid approaches have pushed the boundaries even further, allowing unprecedented control over various aspects of synthesized speech, including voice style, emotion, and even background noise.[37]

Generative Adversarial Networks (GANs) have also made significant contributions to speech synthesis, particularly in the vocoder stage of the text-to-speech pipeline. The vocoder is responsible for converting acoustic features into audio waveforms. GAN-based vocoders like MelGAN[38] and HiFi-GAN[39] have demonstrated the ability to produce high-quality speech waveforms in real-time, addressing some of the speed limitations of earlier models like WaveNet. These GAN-based vocoders have become popular components in many state-of-the-art text-to-speech systems, offering a balance between audio quality and generation speed while working alongside other models that handle text analysis and acoustic feature prediction.

Another notable advancement came with Google Research's AudioLM (2022), which applies language modeling techniques to audio generation.[40] AudioLM represents a novel approach that can generate high-quality, coherent audio continuations given an audio prompt.[41] Unlike previous models that focus solely on speech, AudioLM can handle various audio types, including speech,

[35] *NVIDIA Tacotron 2*, GitHub, https://github.com/NVIDIA/tacotron2 (last visited Aug. 22, 2024).

[36] Yi Ren et al., *FastSpeech 2: Fast and High-Quality End-to-End Text to Speech*, arXiv, https://arxiv.org/abs/2006.04558(last visited Aug. 22, 2024).

[37] *Introducing Voicebox: The First Generative AI Model for Speech to Generalize Across Tasks with State-of-the-Art Performance*, Meta AI, https://ai.meta.com/blog/voicebox-generative-ai-model-speech/ (last visited Aug. 22, 2024).

[38] Kundan Kumar et al., *MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis*, arXiv preprint arXiv:1910.06711 (2019).

[39] Jungil Kong, Jaehyeon Kim & Jaekyoung Bae, *HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis*, in *Proceedings of the 33rd Advances in Neural Information Processing Systems* (2020).

[40] Zalán Borsos et al., *AudioLM: A Language Modeling Approach to Audio Generation*, Google Research, https://google-research.github.io/seanet/audiolm/examples/ (last visited Aug. 22, 2024).

[41] Ibid.

music, and sound effects.[42] This technology demonstrates the potential for more versatile and context-aware voice cloning and audio generation systems, further blurring the line between human and AI-generated audio content.

In March 2024, OpenAI introduced a new text-to-speech model as part of ChatGPT's voice conversation feature.[43] This model represents a significant advancement in voice AI, capable of generating natural-sounding speech in multiple languages and accents from as little as 15 seconds of sample audio. It can also produce speech with varying emotions and styles, further blurring the line between AI-generated and human speech. This development showcases the rapid progress in few-shot learning techniques and the increasing sophistication of voice cloning technologies, pointing towards a future where AI voices are nearly indistinguishable from human ones.

2.3. Voice AI: State of the Art & Technological Challenges

Modern voice cloning technology has made significant strides, leveraging advanced AI techniques to produce increasingly realistic synthetic voices. Today's voice cloning systems can create relatively accurate voice models from small samples of recorded speech, often requiring as little as a few seconds of audio.

This efficiency is made possible by powerful machine learning algorithms and the availability of vast computational resources. More specifically, the capabilities of voice cloning technologies depend on extensive data sets such as LibriTTS[44] and Common Voice[45] that enable AI models to learn from thousands of speakers, improving overall performance and versatility. Advanced "Few-Shot" and "Zero-Shot Learning" techniques allow AI models to clone voices they have never encountered before with minimal reference material.[46] For instance, NVIDIA's RAD-TTS (2021) can generate natural speech in a target voice after hearing only a few reference utterances.[47]

---

[42] Ibid.

[43] *Navigating the Challenges and Opportunities of Synthetic Voices*, OpenAI (Mar. 29, 2024), https://openai.com/index/navigating-the-challenges-and-opportunities-of-synthetic-voices/ (last visited Aug. 22, 2024).

[44] Heiga Zen et al., *LibriTTS: A Corpus Derived from LibriSpeech for Text-to-Speech*, arXiv, https://arxiv.org/abs/1904.02882 (last visited Aug. 22, 2024).
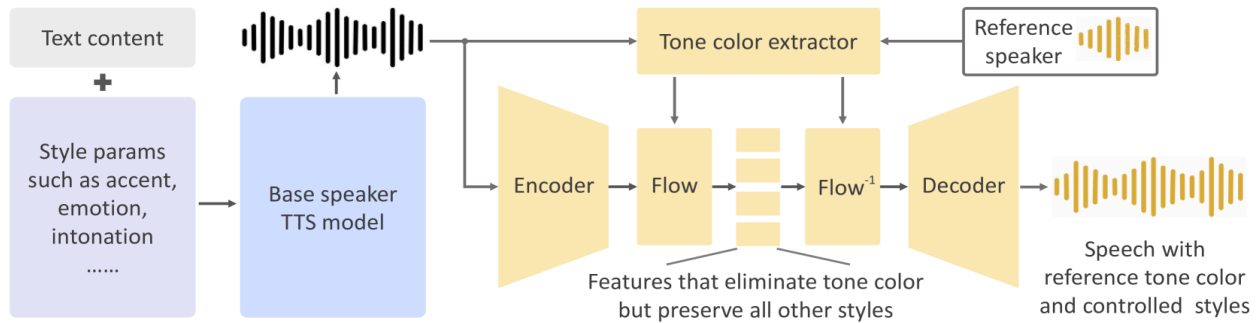
[45] *Common Voice Datasets*, Mozilla, https://commonvoice.mozilla.org/en/datasets (last visited Aug. 22, 2024).

[46] Haitong Zhang & Yue Lin, *Improve Few-Shot Voice Cloning Using Multi-Modal Learning*, arXiv, https://arxiv.org/abs/2203.09708 (last visited Aug. 22, 2024) and Zengyi Qin, *OpenVoice: Versatile Instant Voice Cloning*, arXiv:2312.01479v4 [cs.SD], https://arxiv.org/html/2312.01479v4 (Dec. 21, 2023).

[47] Maggie Zhang & Grzegorz Karch, *Generate Natural Sounding Speech from Text in Real-Time*, https://developer.nvidia.com/blog/generate-natural-sounding-speech-from-text-in-real-time/ (last visited Aug. 22,

The image below provides a visual representation of how advanced Voice AI technologies, such as the OpenVoice framework, function.

Figure 1. Illustration of the OpenVoice Framework.[48]

Despite these recent advancements in extensive data processing capabilities, significant technological challenges persist. The first challenge relates to accurate lip-syncing with video. Achieving precise synchronization between generated audio and the visual features of virtual characters remains challenging.[49] This challenge is complicated by variations in speaking rates and styles across different speakers and languages. Solving this issue is crucial for applications in animation, virtual reality, and video production.

The second challenge relates to the huge number of languages: while voice cloning technology has progressed significantly for major languages like English and Mandarin, support for less common languages lag behind. Expanding language coverage is especially important for countries with smaller populations. Meeting this challenge requires (a) the development of multilingual datasets that capture a diverse range of phonetic inventories and prosodic patterns, (b) the creation of models capable of handling the unique characteristics of multiple languages

2024); Fahima Khanam et al., *Text to Speech Synthesis: A Systematic Review, Deep Learning Based Architecture and Future Research Direction*, 13 J. Adv. Inf. Technol. 398, 398–412 (Oct. 2022).

[48] Zengyi Qin et al., *OpenVoice: Versatile Instant Voice Cloning*, arXiv, https://arxiv.org/html/2312.01479v4 (last visited Aug. 22, 2024).

[49] Chen Chen et al., *Enhancing Zero-Shot Text-to-Speech Synthesis with Human Feedback*, arXiv, https://arxiv.org/html/2406.00654v1 (last visited Aug. 22, 2024).

simultaneously, and (c) adaptation techniques to transfer knowledge from high-resource languages to low-resource ones.[50]

The third challenge for the development of voice AI technologies relates to real-time performance on consumer hardware. Achieving low latency, real-time voice cloning is crucial for interactive applications such as video conferencing, live events, and gaming. This challenge involves balancing model complexity with latency constraints, optimizing inference speed without sacrificing voice quality, and developing efficient streaming architectures that can process audio incrementally.[51]

Addressing these challenges will be crucial for unlocking the full potential of voice AI and voice cloning technology. As researchers continue to innovate, we can expect further improvements in the naturalness, versatility, and applicability of voice cloning systems across various domains. However, the trade-off is that diffusion models are computationally expensive to run, which is likely one reason why OpenAI has been cautious about releasing this technology widely. There are also important concerns around potential misuse that need to be carefully considered.

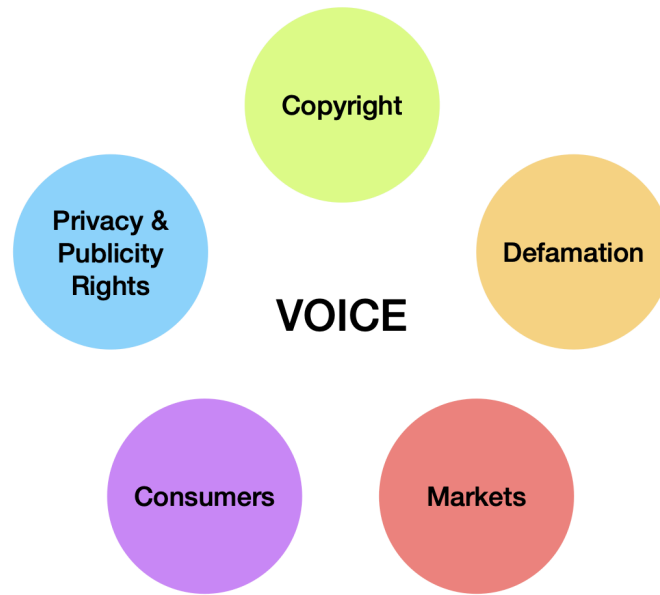# 3. The Failure of Existing Regulatory Frameworks

As an initial step, it is essential to identify the existing legal and ethical rules and principles that underpin discussions on the deployment of voice AI. We argue that the current normative rules and justifications, such as facilitating creativity, ensuring consumer protection, and regulating fair competition in the market, are primarily focused on protecting third parties, whether in the public interest or to ensure healthy markets, rather than the person whose voice is being used. Our aim is to show that these frameworks of regulation do not address the protection/misuse of an individual's voice, creating significant problems and uncertainties.

By exploring the selected six regulatory domains, we offer not just a description but a critique of the current legal and regulatory framework, demonstrating its inadequacy in addressing the novel challenges posed by voice cloning technologies in the age of generative AI. While the existing regulations in the U.S. are well-designed to protect consumers and the public interest and to promote vibrant competition, they fall short of effectively addressing the unique issues related to voice AI. We will introduce the concept that these regulatory areas do not protect voices in any meaningful way; they orbit the issue without directly addressing it.

---

[50] Ye Jia & Michelle Tadmor Ramanovich, *High-Quality, Robust and Responsible Direct Speech-to-Speech Translation*, https://research.google/blog/high-quality-robust-and-responsible-direct-speech-to-speech-translation/ (last visited Aug. 22, 2024).

[51] Mars5-TTS by CAMB-AI claims to have addressed the latency issues and only require 5s of sample data to imitate someone's speech. Its adoption is yet to be seen.

Figure 2: The Limits of the Multiple Regulatory Domains

The section is organized to highlight the most plausible legal grounds that individuals might use in cases of unlawful or permissionless use of their voice, illustrating the predictability and remedies available under existing laws. In this section, we focus specifically on copyright law, the protection of privacy and personality rights, and the protection of an individual's reputation, consumers, and markets. We also address an overlooked area of the so-called "industry agreements" between content creators and information intermediaries in such creative industries as Hollywood. This approach not only underscores the gaps in current regulations but also paves the way for discussing innovative legal solutions. By identifying these deficiencies, we aim to lay the groundwork for necessary and desirable improvements in lawmaking, thereby advocating for a robust legal framework that can effectively address the complexities introduced by voice AI technologies.

Before we go into details, it is important to observe that historically, an individual's voice has not been protected by law at all. In one sense, this is not surprising because it was not previously an issue. The protection of voice has only become important with the advancement of technological innovations.

3.1. Protecting Creativity

The public availability of generative AI tools brought to the attention various attempts by amateur and professional creators to explore the capabilities and limits of those new AI tools. In February 2023, one of the world's most renowned dance music producers, French DJ David Guetta, shared a clip on Instagram from one of his latest concerts.[52] The clip showcased a segment where the voice, closely reminiscent of Eminem, declares, "This is the future rave sound; I'm getting off the underground," set against the backdrop of pulsating dance music. This innovation was a striking example of Future Rave (FR), a genre Guetta has pioneered alongside Danish DJ Morten Breum.

What set this particular instance apart was Guetta's admission of employing generative AI to create this segment. In a candid Instagram post, Guetta explained that the Eminem-style verse was generated using AI tools—one to create lyrics in the rapper's iconic style and another to replicate his distinctive voice. This blend of content, which Guetta was able to create with generative AI tools, resulted in an electrifying concert experience, as he noted, "People went nuts."

However, despite the enthusiastic reception, Guetta acknowledged the complex legal and ethical dimensions of using AI in this manner. Out of respect for Eminem and the murky waters of copyright law in the AI age, he decided against releasing the mix publicly.[53] In another interview with the BBC, he stated that, "I'm sure the future of music is in AI. For sure. There's no doubt. But as a tool. … Nothing is going to replace taste. What defines an artist is that you have a certain taste, you have a certain type of emotion you want to express, and you're going to use all the modern instruments to do that."[54]

David Guetta's experimentation with AI tools is by far not the only instance that captured public attention. In April 2023, an incognito creator known as Ghostwriter released a song titled "Heart on My Sleeve" which mimicked the voices of Drake and The Weeknd.[55] The song was created without the consent from The Weeknd and Drake. It quickly gained massive popularity and went viral online, but the initial post was soon removed from streaming platforms due to copyright and moral concerns.[56]

---

[52] See https://www.instagram.com/p/CoNqQuFqIHZ/ (last visited Feb. 12, 2023).

[53] Sam Roche, *David Guetta: "If You Have Terrible Taste, Your Music Is Still Gonna Be Terrible, Even with AI"*, MusicTech (July 24, 2023), https://musictech.com/news/david-guetta-on-ai/ (last visited Aug. 30, 2023).

[54] *David Guetta Says the Future of Music Is in AI*, BBC (Feb. 13, 2023), https://www.bbc.com/news/entertainment-arts-64624525 (last visited Aug. 30, 2023).

[55] Ghostwriter - Heart on My Sleeve Drake ft. The Weeknd (Remastered AI Song), https://www.youtube.com/watch?v=rQssjhX31Z0 (last visited Aug. 22, 2024).

[56] Kristin Robinson, *Ghostwriter, the Mastermind Behind the Viral Drake AI Song, Speaks for the First Time*, Billboard (Oct. 11, 2023),

Another instance involved Lady Gaga fans, who, in response to the prolonged wait for new official releases, took matters into their own hands by creating and sharing songs that stimulated her voice on a YouTube channel. This trend has been particularly evident on the channel Originals By Little Monsters, which has released AI versions of tracks from the anticipated but unreleased "ARTPOP Act II" album, including songs like "TEA"[57] and a remix of "Partynauseous."[58] By employing various generative AI tools, fans generated covers and original songs that sound remarkably like Lady Gaga, filling the void left by the absence of new music. While some fans find these AI-generated tracks entertaining and a creative way to keep the community engaged, others express concerns about the ethical implications and the potential for misuse of such technology.[59]

These incidents prompted music industry giant Universal Music Group ("UMG") to send urgent letters to streaming platforms (including Spotify and Apple Music) requesting them to block AI companies from collecting melodies and lyrics from copyrighted songs to create new AI-generated songs, citing a "moral and commercial responsibility" to prevent unauthorized use of artists' voices.[60] Similarly, Sony made a public announcement that it objects to using any of its content to train AI models.[61]

Could copyright laws be invoked as a shield or as a sword in situations where a third party collaborates with AI tools to create a deepfake version of someone's voice? The answer is not straightforward. Copyright laws emerged as a tool for artists and creators to protect their creative works, such as music compositions, sound recordings, synchronization rights, related digital rights, and rights to the masters, which are usually owned by record labels. To protect the interests of creators, copyright law grants exclusive rights, enabling them to control how their works are recorded, reproduced, and distributed. Nowadays, when a music artist records a song, her vocal performance is fixed in a digital audio file (the "master"). In this context, copyright law

https://www.billboard.com/music/pop/ghostwriter-heart-on-my-sleeve-drake-ai-grammy-exclusive-interview-1235434099/.

[57] Originals by Little Monsters, Lady Gaga - TEA (AI Original),
https://www.youtube.com/watch?v=oO8lBKv7CV0&t=46s (last visited Aug. 22, 2024).

[58] Originals by Little Monsters, Lady Gaga - Partynauseous (AI Original),
https://www.youtube.com/watch?v=6NZnzKnZ5pw (last visited Aug. 22, 2024).

[59] See e.g., Taylor Alexis Heady, *Fans Are Using AI to Make Their Favorite Pop Stars Sing Whatever They Want*, PopCrush (Jan. 23, 2023), https://popcrush.com/ai-artists-sing-any-song-ariana-grande-lady-gaga/#.

[60] Amanda Silberling, *From DrakeGPT to Infinite Grimes, AI-Generated Music Strikes a Chord*, TechCrunch, https://techcrunch.com/2023/04/26/grimes-ai-generated-drake-music-legal-issues/ (last visited July 18, 2024).

[61] Daniel Tencer, *Sony Music Sends Letters to 700 AI, Music Streaming Companies Declaring It's 'Opting Out' of AI Training*, Music Business Worldwide, https://www.musicbusinessworldwide.com/sony-music-sends-letters-to-700-ai-music-streaming-companies-declaring-its-opting-out-of-ai-training1/ (last visited Aug. 22, 2024).

protects this "master," which is the tangible fixation of the artist's voice as an original work of authorship.[62] However, copyright law does not protect the voice of the performer itself. For example, if an artist is performing a song that is being recorded, copyright law protection will extend to the vocal performance as expressed by the artist's distinct tone, style, and delivery if it is fixed in a tangible medium.

Conceptually, copyright law can protect a creator's voice against illegal acts of third parties utilizing AI voice cloning tools. The reproduction and imitation of a creator's voice without permission can constitute an infringement of copyright. The exclusive right to reproduce the copyrighted work, which includes vocal performances, is a core aspect of copyright law. Therefore, unauthorized voice cloning or imitation that involves the reproduction of the creator's vocal performance may be considered a violation of their copyright. However, this protection extends to the actual sounds fixed in the recording, and the law aims to prevent unauthorized duplication or imitation of these sounds without the creator's consent. In other words, copyright laws in general, and the right of reproduction in particular, seem to protect only specific instances of unlawful copying of expressions that are fixed, yet copyright does not protect against the use of an individual's voice per se or likeness.

In 2023, the U.S. Copyright Office launched a study to explore the impact of AI on Copyright.[63] During the several months of the comment period, more than 10,000 comments were submitted by various stakeholders from all around the world who expressed their positions on a wide array of topics, including AI-powered voice cloning.[64] In Section 4 of this paper, we will reflect on some of the key findings from the stakeholder analysis. For now, it suffices to say that we need more clarity on how copyright laws will address AI technologies powering voice cloning technology.

To sum up, contemporary copyright laws do not provide an answer about the legality of AI-powered voice cloning technologies. *De lege lata*, copyright law does not protect a human's voice, nor does copyright protection extend to the style of speaking or singing. More radically, while copyright law presents itself as presenting creators and creativity, the way how copyright laws (at least in the US) have evolved during the past century, copyright ended up protecting information intermediaries. It would be great to have more certainty on Voice AI

---

[62] *Melodies and Machines: Copyright Challenges with AI Music*, JD Supra, https://www.jdsupra.com/legalnews/melodies-and-machines-copyright-1294672/#_ftn1 (last visited Aug. 22, 2024).
[63] See https://www.copyright.gov/policy/artificial-intelligence/
[64] See: https://www.regulations.gov/docket/COLC-2023-0006/comments

3.2. Protecting Privacy

In the United States, privacy rights are a relatively recent phenomenon, having evolved significantly over the past century.[65] These rights emerged before the formal recognition of publicity rights, which protect individuals from the unauthorized commercial use of their likeness.[66] Privacy serves as the foundational basis for the right of publicity, underscoring the importance of safeguarding personal space and autonomy in the order of things. The concept of privacy is centered on protecting the intimate aspects of individuals' lives from intrusion, thereby providing the essential groundwork upon which publicity rights are built. This framework becomes particularly relevant in the age of Voice AI, where the lines between personal privacy and public persona are increasingly blurred.

In the following sections, we outline how the right of publicity has evolved in specific US states. Focusing on the laws of California, New York, and Tennessee, we illustrate some of the emerging regulations regarding the protection of voice in the Age of AI. These developments serve as a solid starting point for a broader discussion about the new social contract between humans and Voice AI technologies, which we will explore in Section 4 of this paper.

3.2.1. US Common Law

As mentioned above, legal protections for an individual's voice have evolved from broader rights of privacy and publicity. The right of publicity is a modern, twentieth-century invention of the law.[67] Hence, there are not that many cases pertaining to the protection of voice from unauthorized exploitation. At the U.S. common law, mere vocal imitations were not deemed to infringe on a celebrity's privacy or publicity rights until 1988.[68] Moreover, such imitation has been held not to constitute unfair competition or passing off. For the purposes of this discussion, we would like to highlight two landmark cases in which the American courts have applied the right of publicity to protect voices that could be relevant to generative AI contexts.

---

[65] Samuel Warren & Louis Brandeis, *The Right to Privacy*, 4 Harv. L. Rev. 193 (1890); Dorothy J. Glancy, *The Invention of the Right to Privacy*, 21 Ariz. L. Rev. 1 (1979); Urs Gasser, *Recoding Privacy Law: Reflections on the Future Relationship Among Law, Technology, and Privacy*, 130 Harv. L. Rev. 171 (2016); Jeffrey Bellin, *Pure Privacy*, 115 Nw. U. L. Rev. 463 (2021).

[66] Art Law Deskbook, Ch. 3.02.

[67] Marshall Leaffer, *The Right of Publicity: A Comparative Perspective*, 70 Alb. L. Rev. 1357 (2007); Art Law Deskbook, Ch. 3.02.

[68] Business Torts § 33.11. See *Sinatra v. Goodyear Tire & Rubber Co.*, 435 F.2d 711 (9th Cir. 1970), cert. denied, 402 U.S. 906 (1971) (applying California law); see also *Davis v. Trans World Airlines*, 297 F. Supp. 1145 (C.D. Cal. 1969) (applying California law).

In the US, a person's name, voice, likeness, or persona could be protected under the body of rules known as "the right of publicity". Artists and celebrities often communicate through symbolic speech, which is geared towards one instrumental purpose – to stand out. Celebrities and sports figures oftentimes are associated with such traits as creativity, curiosity, bravery, hubris, debauchery, sophistication, or many other aspects of the human condition.[69] Since achieving celebrity status usually requires time, effort, and resources, it is natural that celebrities are keen to protect their names, voices, or likeness.

The right of publicity covers protections of a person's name, image, likeness, persona, and often their voice, as well as other distinctive characteristics, from unauthorized commercial exploitation by others.[70] The right has been construed very broadly. As a matter of principle, the right of publicity in the US encompasses basically anything that can be exploited by the celebrity herself or can be associated with the celebrity.[71]

The first case is *Midler v. Ford Motor Co.*,[72] where the court granted protection to the sound of Bette Midler's singing voice. Bette Midler is an American singer, actress, comedian, and author best known for her powerful voice and dynamic performances. Some of her most famous songs include "The Rose," "Wind Beneath My Wings," and "From a Distance." She has won Grammy Awards, Golden Globes, and a Tony Award. This case revolved around the unauthorized use of Midler's distinctive voice in a commercial by Ford Motor Company. The company had licensed the song "Do You Wanna Dance" and had hired a sound-alike to imitate Midler's voice.

In 1986, Midler filed a complaint after an advertising agency created a television commercial using a sound-alike of her voice. The district court initially granted the defendant's motion for summary judgment, stating that Midler had no protectable rights under California law. However, the Ninth Circuit reversed this decision, holding that when a distinctive voice of a professional singer is widely known and deliberately imitated to sell a product, the sellers have appropriated what is not theirs and have committed a tort in California.[73]

The Midler standard was further confirmed and clarified by the Ninth Circuit in *Waits v. Frito-Lay, Inc.*[74] The case involved an American singer Tom Waits who sued Frito-Lay (a

---

[69] Art Law Deskbook, Ch. 3.00.

[70] Art Law Deskbook, Ch. 3.01.

[71] Art Law Deskbook, Ch. 3.01.

[72] *Midler v. Ford Motor Co.*, 849 F.2d 460 (9th Cir. 1988).

[73] *Midler v. Ford Motor Co.*, 849 F.2d 460, 463 (9th Cir. 1988) (declining to find unfair competition or violation of California's privacy statute). For a public comment, see Note, *"Do You Want to Dance" Around the Law? Learn the Latest Steps from the Ninth Circuit in Midler v. Ford Motor Co.*, 23 Loy. L.A. L. Rev. 601 (1990); Note, *Voicing Concern: An Overview of the Current Law Protecting Singers' Voices*, 40 Syracuse L. Rev. 1255 (1989); Comment, *Whose Voice Is It Anyway? Midler v. Ford Motor Company*, 8 Cardozo Arts & Ent. L.J. 201 (1989).

[74] *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093, 1096, 1098 (9th Cir. 1992), cert. denied, 506 U.S. 1080 (1993).

subsidiary of PepsiCo that manufactures, markets, and sells a wide variety of snack foods) and its advertising agency for imitating his distinctive (described as "raspy" and "gravelly") singing voice in a commercial.[75] Frito-Lay knew that Waits would not perform for the commercial because of his public stance against musicians in advertising. Therefore, it contracted another singer who produced an imitation of Waits's voice in the advertisement's music (which was a close copy of Waits's song "Step Right Up"). At the trial, the jury found that Frito-Lay had violated Waits's right of publicity and awarded Waits $375,000 in compensatory damages and $2 million in punitive damages.[76]

Waits's jury award was upheld by the Ninth Circuit on appeal. The court accepted Midler's three-part test for establishing a sound-alike cause of action, which required that the voice be (a) distinctive, (b) widely known, and (c) deliberately imitated for commercial use as the correct framework.[77] Moreover, the Ninth Circuit held that a performer's "style" of singing was not protected and limited the right of publicity to protect only the performer's voice.[78] In interpreting the meaning "distinctiveness" and "widely known", the court noted that "a voice is distinctive if it is distinguishable from the voice of other singers … if it has particular qualities or characteristics that identify it with a particular singer.[79]

3.2.2. California

In the U.S., some states have adopted special right of publicity statutes and entrench the possibility to control the commercial exploitation of a person's voice.[80] For example, the California Civil Code includes "voice" in its definition of the right of publicity.[81] The California Legislative Counsel has confirmed that there are strong public policy reasons to permit imitations, as long as they are not passed off as originals, in order to promote competition.[82] Several subsequent court cases in California clarified that "voice" protections only encompass the actual voice of a person and do not include (synthetic robot) imitations.[83]

---

[75] Business Torts § 33.11.

[76] *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093, 1101 (9th Cir. 1992), cert. denied, 506 U.S. 1080 (1993).

[77] *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093, 1100–01 (9th Cir. 1992), cert. denied, 506 U.S. 1080 (1993).

[78] *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093, 1100 (9th Cir. 1992), cert. denied, 506 U.S. 1080 (1993).

[79] *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093, 1101 (9th Cir. 1992), cert. denied, 506 U.S. 1080 (1993).

[80] See, e.g., Cal. Civ. Code § 3344; Ohio Rev. Code Ann. §§ 2741.01(A), 2741.01(A).

[81] § 3344 of Cal. Civ. Code was amended in 1984 to specifically add "voice" as a protected element under the right of publicity. See Business Torts § 33.11.

[82] Letter from Bion M. Gregory, Legislative Counsel of California, to Senator William Campbell (Dec. 14, 1984).

[83] See *White v. Samsung Elec. Am., Inc.*, 989 F.2d 1512 (9th Cir.), cert. denied, 508 U.S. 951 (1993) ("Without deciding for all purposes when a caricature or impressionistic resemblance might become a 'likeness,' we agree with the district court that the robot at issue here was not White's 'likeness' within the meaning of section 3344.").

3.2.3. New York

New York Civil Rights Law (CRL) §50 and §51 provides legal grounds for cases of illegal use of a celebrity's voice.[84] The NY CRL prohibits the use of a person's name, portrait, picture, or voice for advertising or trade purposes without the person's written consent.[85] A complaint based on §51 must show that: (1) the defendant used the plaintiff's name, portrait, picture, or voice, (2) for purposes of trade or advertising, and (3) without his or her written consent.[86] If a violation is established, the plaintiff can maintain an equitable action to prevent and restrain the use thereof and recover damages for any injuries sustained by reason of such use.[87] If the defendant knowingly used the plaintiff's name, portrait, picture, or voice in such a manner as is forbidden or declared to be unlawful by section fifty of this article, the jury, in its discretion, may award exemplary damages.[88] The protections enshrined in the NY CRL are limited to nonconsensual commercial appropriations and do not apply to reports of newsworthy events or matters of public interest.[89]

3.2.4. Tennessee

Some U.S. states, such as Tennessee, are now choosing to legislate specifically to protect musicians and other celebrities against AI deep fakes. The newly adopted Ensuring Likeness Voice and Image Security (ELVIS) Act provides protection for an individual's likeness and extends the right of publicity to their voice.[90] Under this law, "voice" is defined broadly and includes both the use of an actual voice and a simulation thereof.

---

[84] §50 provides "A person, firm or corporation that uses for advertising purposes, or for the purposes of trade, the name, portrait, picture, likeness, or voice of any living person without having first obtained the written consent of such person, or if a minor of such minor's parent or guardian, is guilty of a misdemeanor."

[85] *Spicer v. Conde Nast Entertainment, LLC*, 2014 N.Y. Misc. LEXIS 5156; *Talley v. Moss*, 2009 N.Y. Misc. LEXIS 5131; *Nussenzweig v. DiCorcia*, 11 Misc. 3d 1051(A) (N.Y. Sup. Ct. 2006).

[86] *Spicer v. Conde Nast Entertainment, LLC*, 2014 N.Y. Misc. LEXIS 5156; *Nussenzweig v. DiCorcia*, 11 Misc. 3d 1051(A) (N.Y. Sup. Ct. 2006); *Alvidrez v. Roberto Coin, Inc.*, 6 Misc. 3d 742 (N.Y. Sup. Ct. 2004).

[87] *Talley v. Moss*, 2009 N.Y. Misc. LEXIS 5131; *Alvidrez v. Roberto Coin, Inc.*, 6 Misc. 3d 742 (N.Y. Sup. Ct. 2004); *Leviston v. Jackson*, 43 Misc. 3d 229 (N.Y. Sup. Ct. 2014).

[88] *Talley v. Moss*, 2009 N.Y. Misc. LEXIS 5131

[89] *Nussenzweig v. DiCorcia*, 11 Misc. 3d 1051(A) (N.Y. Sup. Ct. 2006); *Leviston v. Jackson*, 43 Misc. 3d 229 (N.Y. Sup. Ct. 2014); *Porco v. Lifetime Entertainment Servs., LLC*, 147 A.D.3d 1253 (N.Y. App. Div. 2017).

[90] See, e.g., Catie Lane Bailey, Tanisha Pinkins & Lauren Caverly Pratt, *First-of-Its-Kind AI Law Addresses Deep Fakes and Voice Clones*, Holland & Knight LLP, https://www.hklaw.com/en/insights/publications/2024/04/first-of-its-kind-ai-law-addresses-deep-fakes-and-voice-clones (last visited Aug. 22, 2024).

The ELVIS Act, which comes into effect on July 1, 2024, protects against any unauthorized commercial use of an individual's voice that is "readily identifiable." More importantly, the ELVIS Act not only protects against the voice cloning of famous artists but also of podcasters and voice actors at all levels of fame.[91] The ELVIS Act explicitly provides that voice cloning is only possible with the individual's consent. Failure to obtain consent is supported by a private right of action against AI companies and criminal actions.

3.3. Protecting Reputation

Defamation law, fundamentally aimed at protecting one's reputation from false statements, presents unique challenges when applied to Voice AI technologies. The core of defamation is the protection against falsehoods that harm an individual's good name, but it does not extend to safeguarding the intrinsic qualities of a person's voice.[92] Defamation could provide grounds for action if the cloned voice is used to propagate false statements that damage an individual's reputation. However, such an application does not capture the full scope of the intrusion, as it overlooks the distinct and paramount issue of voice misappropriation. Thus, while defamation rules offer redress for reputational harm, they do not directly address the unauthorized use or replication of one's voice, which represents a significant gap in the context of Voice AI.

Protecting one's voice from unauthorized use in the age of AI necessitates novel legal frameworks beyond defamation, as the harm involved is not solely reputational but also deeply personal and expressive. The unlawful use of an individual's voice constitutes a violation of their personal identity, requiring more comprehensive legal protections. Therefore, while defamation rules are relevant, they are insufficient on their own, underscoring the need for new legal paradigms to address the challenges posed by Voice AI.

3.4. Protecting Consumers

Another set of relevant regulations proceed from the perspective of protecting consumers from misleading representations about products of services.

---

[91] Ibid.

[92] John B. Lewis & Bruce L. Ottley, *New York Times v. Sullivan at 60: Where Does Defamation Law Go Now?*, 73 DePaul L. Rev. 995 (2024), https://via.library.depaul.edu/law-review/vol73/iss4/3; Yonathan A. Arbel & Murat Mungan, *The Case Against Expanding Defamation Law*, 71 Ala. L. Rev. 453 (2020); Alexandra M. Gutierrez, *The Case for a Federal Defamation Regime*, Yale L.J. F. (Sept. 15, 2021), https://www.yalelawjournal.org/forum/the-case-for-a-federal-defamation-regime.

3.4.1. The Lanham Act: A Framework for Trademarks to Mitigate Confusion

In the US, the Lanham Act of 1946, also known as the Trademark Act, provides a national system for trademark registration and protects trademark owners against the unauthorized use of their marks.[93] It contains provisions for trademark infringement, false advertising, and false endorsement, ensuring that consumers are not misled about the origins of goods and services. The act also grants trademark holders the right to sue for damages and obtain injunctive relief against those who violate their trademark rights.

The Lanham Act is particularly relevant in the context of Voice AI technologies because it provides robust protection against the misuse of an individual's voice in commercial contexts. Any misuse of an individual's voice is treated as a false endorsement of a product or a service.[94] Section 43(a) of the Lanham Act, codified at 15 U.S.C. § 1125(a), prohibits the use of any symbol or device in commerce that is likely to cause confusion or deceive consumers regarding a person's affiliation, connection, or association with another person, or as to the origin, sponsorship, or approval of goods, services, or commercial activities by another person.[95]

Specifically, the Act makes actionable the deceptive and misleading use of marks, including an individual's distinctive voice, which can be considered a "symbol or device" under the Lanham Act. The courts have recognized that a false endorsement claim can be based on the unauthorized use of a celebrity's identity, including vocal imitation, which is likely to confuse consumers as to the plaintiff's sponsorship or approval of the product.[96] This interpretation is supported by case law, which has consistently held that the unauthorized imitation of a celebrity's distinctive voice for commercial gain without their approval constitutes a false endorsement under the Lanham Act.[97]

3.4.2. Voice AI, Robocalls & Telemarketing

The use of software to generate automated messages and apply them in telemarketing has rather deep roots in the U.S. In response to growing consumer concerns about intrusive telephone and

---

[93] *The Lanham Act*, Pub. L. No. 79-489, 60 Stat. 427 (1946) (codified as amended at 15 U.S.C. § 1051 et seq.).

[94] *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093 (9th Cir. 1992); *Abdul-Jabbar v. GMC*, 85 F.3d 407 (9th Cir. 1996); and *Butler v. Target Corp.*, 323 F. Supp. 2d 1052 (C.D. Cal. 2004).

[95] *Dryer v. NFL*, 814 F.3d 938 (8th Cir. 2016); *Facenda v. NFL Films, Inc.*, 542 F.3d 1007 (3d Cir. 2008); and *Roberts v. Bliss*, 229 F. Supp. 3d 240 (S.D.N.Y. 2017).

[96] *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093 (9th Cir. 1992); *Abdul-Jabbar v. GMC*, 85 F.3d 407 (9th Cir. 1996); and *Butler v. Target Corp.*, 323 F. Supp. 2d 1052 (C.D. Cal. 2004).

[97] *Waits v. Frito-Lay, Inc.*, 978 F.2d 1093 (9th Cir. 1992); *McGillvary v. Vonkurnatowski*, 2024 U.S. Dist. LEXIS 65993; *Abdul-Jabbar v. GMC*, 85 F.3d 407 (9th Cir. 1996).

fax solicitations, Congress enacted two significant statutes aimed at telemarketing. In 1991, it passed the Telephone Consumer Protection Act (TCPA)[98] which mandated the Federal Communications Commission (FCC) to create rules on telemarketing by phone and fax. Three years later, in 1994, Congress enacted the Telemarketing and Consumer Fraud and Abuse Prevention Act[99], directing the Federal Trade Commission (FTC) to establish rules targeting deceptive and abusive telemarketing practices, including credit card laundering.

The FCC regime focuses on telephony solicitation laws enforced by the FCC alongside those by the FTC. The TCPA specifies numerous prohibitions and requirements for telemarketing activities, including the use of prerecorded messages, auto-dialing, unsolicited fax messages, and calls to residences. It also mandates the creation and maintenance of a national do-not-call (DNC) list.[100] The TCPA authorizes the FCC to prescribe regulations on the use of automated phone equipment and provides a private right of action in state court for violations, allowing for injunctive relief and damages.[101]

Additionally, the TCPA permits the establishment of a single national database for residential subscribers who object to receiving phone solicitations. This list must be available for purchase, and a private right of action is granted to individuals who receive multiple calls from the same entity within a 12-month period in violation of these regulations.[102] The statute also covers unlawful conduct involving fax machines, artificial or prerecorded voice, automatic dialing systems, and caller ID information manipulation. TCPA applies to any phone call initiated using an artificial or prerecorded voice message, even if a live agent is present on the call.[103] Based on the TCPA, consumers could invoke theories of (lack of) consent and violation of privacy to prevent the use of voice cloning technologies to target consumers. Violators face forfeiture penalties and potential criminal fines, and state attorneys general can bring actions on behalf of their residents.

This statute and its authorized rules generally do not preempt state laws that impose stricter requirements or prohibit unsolicited ads, automated dialing systems, or phone solicitations.[104] State attorneys general are empowered to bring civil actions in federal court to address patterns of calls violating the statute, seeking injunctions and monetary damages. The FCC is required to

---

[98] *Telephone Consumer Protection Act*, Pub. L. No. 102-242, 47 U.S.C. § 227.

[99] 15 U.S.C. § 6101 et seq.

[100] See https://www.donotcall.gov/.

[101] 5 COMPUTER LAW: A Guide to Cyberlaw and Data Privacy § 28.07 (2024).

[102] The amount is limited to $500, however, at the court's discretion, damages may be increased to not more than three times this amount for a willful or knowing violation.

[103] See FCC's Soundboard decision.

[104] 5 COMPUTER LAW: A Guide to Cyberlaw and Data Privacy § 28.07 (2024).

report to Congress annually on the enforcement of TCPA provisions related to unsolicited fax advertisements.

More recently, on December 30, 2019, in response to public outcry over robocalls, the President signed the TRACED Act into law.[105] This statute mandates the FCC to require voice service providers to implement the STIR/SHAKEN authentication framework and other effective call authentication measures within 18 months. The technology must be implemented without cost to consumers and small businesses. Besides, the FCC must issue best practices for call authentication within one year and assess the efficacy of these technologies every three years, revising them as necessary and reporting to Congress. The TRACED Act also requires the FCC to establish rules for blocking voice calls based on call authentication frameworks, providing a safe harbor for inadvertent blocking or misidentification and ensuring calls from compliant areas are not unreasonably blocked.

An interagency working group, led by the Attorney General and FCC Chairman, oversees prosecuting violations of the Communications Act of 1934. Furthermore, the FTC must initiate proceedings to protect against "one-ring scams," and the TCPA forfeiture penalties were expanded, imposing new time limitations on bringing actions.

In 2024, the FTC adopted a ruling that banned unwanted AI-generated robocalls.[106] This ruling classifies calls made with AI-generated voices as "artificial" under the TCPA, thereby prohibiting telemarketing calls that use automatic dialing systems and artificial or prerecorded voice messages without prior consent from recipients. The decision was made in response to concerns about the misuse of AI technology for deceptive purposes, particularly in the context of election misinformation and scams targeting vulnerable individuals.[107]

3.5. Protecting Markets

Next, we consider an antitrust perspective, and a range of regulations aimed at protecting the integrity and operations of the market.

3.5.1. Fair Competition & Antitrust Laws

---

[105] 5 COMPUTER LAW: A Guide to Cyberlaw and Data Privacy § 28.07 (2024).
[106] *FCC Makes AI-Generated Voices for Robocalls Illegal*, FCC, https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal (last visited Aug. 22, 2024).
[107] Cecilia Kang, *FCC Bans AI-Generated Robocalls*, N.Y. Times (Feb. 8, 2024), https://www.nytimes.com/2024/02/08/technology/fcc-ban-ai-robocalls.html.

The fair functioning of the market in the U.S. is ensured by three federal statutes: the Sherman Antitrust Act of 1890, the Clayton Act of 1914, and the Federal Trade Commission Act of 1914. The Sherman Antitrust Act serves as the cornerstone of U.S. antitrust law, prohibiting anticompetitive practices such as monopolization, price fixing, and other trade restraints. Specifically, Section 1 of the Sherman Act prohibits contracts, combinations, and conspiracies that unreasonably restrain interstate and foreign trade, including agreements among competitors to fix prices, rig bids, or divide markets. Section 2 targets monopolization, attempted monopolization, and conspiracies to monopolize, making it illegal for any company to monopolize or attempt to monopolize any part of interstate commerce.

The second pillar of the American Federal law which aims to protect markets is the Clayton Act, which primarily aims to prevent anti-competitive practices that could lead to monopolies or restrain trade.[108] It includes provisions that define antitrust laws, address the consequences of antitrust violations, and outline the scope of permissible legal actions against entities that engage in such practices. The Clayton Act facilitates easier litigation for private parties injured by antitrust violations by allowing prior government antitrust judgments to serve as prima facie evidence in subsequent private lawsuits.[109]

The Federal Trade Commission Act of 1914 is the third pillar of US antitrust laws to protect the markets. The FTC Act bans unfair methods of competition and unfair or deceptive acts affecting commerce. From the procedural perspective, the FTC is significant because it empowers the Federal Trade Commission (FTC) to prevent such anti-competitive behavior. The FTC Act explicitly states that "unfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce, are declared unlawful."[110]

The FTC is vested with the authority to issue cease and desist orders to halt unfair practices. Also, the FTC can require affirmative disclosure of material facts that might influence consumers' decisions.[111] This ensures that businesses operate transparently and do not engage in deceptive practices that could mislead consumers or harm other businesses. Furthermore, the FTC Act empowers the FTC to seek injunctive relief in federal district court to prevent ongoing or imminent unfair practices, thereby providing a robust mechanism to address anti-competitive behavior swiftly and effectively.[112] This broad discretion allows the FTC to intervene in a wide

---

[108] *Clayton Act*, ch. 323, 38 Stat. 730 (1914) (codified as amended at 15 U.S.C. §§ 12–27 and 29 U.S.C. §§ 52–53).
[109] *Union Carbide & Carbon Corp. v. Nisley*, 300 F.2d 561 (10th Cir. 1962).
[110] *Federal Trade Commission Act*, ch. 311, 38 Stat. 717 (1914) (codified as amended at 15 U.S.C. §§ 41–58).
[111] *In the Matter of Peacock Buick, Inc., a Corporation, and Dr. Norman Bernstein and Michael B. Peacock, Individually and as Officers of Said Corporation*, FTC Docket No. XXXX (Year).
[112] *FTC v. Accusearch, Inc.*, 570 F.3d 1187 (10th Cir. 2009).

range of activities to ensure the discontinuance of unlawful practices, promote fair competition, and protect consumers and businesses from deceptive practices.[113]

3.5.2. The FTC's Plans for Voice AI

In the administrative structure of U.S. market governance, the FTC is a key player in shaping fair competition practices across various industries. However, because voice cloning technologies are relatively new, the FTC has not yet had the opportunity to apply federal antitrust laws to voice cloning controversies. Some limited insights can be gleaned from public comments by FTC Chairwoman Lina Khan, who has highlighted the FTC's determination to monitor the development of AI technologies in the U.S.

In one of her recent statements, Lina Khan expressed concerns about the potential misuse of AI technologies, particularly regarding the use of data and control over AI tools.[114] She noted the pressure Big Tech companies place on startups and independent entrepreneurs developing AI technologies. Khan emphasized the apprehensions surrounding access to AI models and the terms of competition, which could lead to discriminatory practices or exclusion, benefiting larger companies at the expense of smaller ones. She also raised issues about AI systems potentially mimicking content producers without proper compensation, supervision, and control. She alluded to the fact that the FTC is aware of numerous risks that emerge when generative AI tools are used to create fraudulent accounts, deep fakes, and voice clones. Khan noted that AI models might violate antitrust laws and that using an artist's or creator's data to produce AI products that replicate their likeness could be "an unfair method of competition."[115]

These concerns reflect a broader apprehension about the ethical implications and fairness of AI technologies, especially in how they handle personal data and the likeness of individuals. The discussions underscore the need for regulatory vigilance and possibly new frameworks to ensure that AI technologies are used in a manner that respects personal rights and promotes fair and open competition.

3.6. Protecting Industry

---

[113]*In the Matter of Peacock Buick, Inc., a Corporation, and Dr. Norman Bernstein and Michael B. Peacock, Individually and as Officers of Said Corporation*, FTC Docket No. XXXX (Year).

[114] Lina Khan, *We Must Regulate A.I. Here's How*, N.Y. Times (May 3, 2023), https://www.nytimes.com/2023/05/03/opinion/ai-lina-khan-ftc-technology.html.

[115] ibid.

In May of 2023, many actors took to the picket lines, launching what would become the longest strike in Tinseltown's history—a grueling 181-day standoff. Indeed, it was the longest work stoppage in Hollywood since the infamous labor battles of 1960. This period quickly became known as the "Hot Strike Summer," and rippled well into the fall, sending shockwaves through the entertainment ecosystem both within the US as well as internationally.[116] Yet, this controversy wasn't just any labor dispute. It was the creative industry's response to a rapidly evolving technological landscape, one where emerging fears about generative AI technologies served as a catalyst for artists' battle for recognition and compensation. The rise of AI pushed the industry past the boiling point. Highly sophisticated machine learning tools now seemed as if they had the capability to replace actors with lifelike AI avatars and swipe the voices of performers with a few lines of code. Writers and directors felt fearful that their unique styles could be replicated by studios and tech giants proprietary generative AI solutions.[117]

The spark that ignited this strike was primarily related to the actors' call for a fair share from the revenue of the Hollywood studios. Yet, it was the sudden emergence of powerful generative AI tools like ChatGPT that gave voice to the unspoken fears and grievances of artists who felt that their creative contributions were undervalued in Hollywood's mind-boggling $130 trillion industry.

In the spring of 2023, the Screen Actors Guild (SAG), the Writers Guild of America (WGA), and the Directors Guild of America (DGA) found themselves at a contractual crossroads when the Alliance of Motion Picture and Television Producers (AMPTP) refused to renegotiate previous agreements. The strike that began with the Writers Guild of America in May gained momentum when SAG members joined in July. One of the central concerns in this AI-infused upheaval was the protection of artists' likenesses from becoming grist for the generative AI mill.

After months of negotiations, in November 2023, the representatives of artists were able to reach agreements with the AMPTP that placed some safeguards against the encroachment of AI. The DGA's contract insisted that AI was not a person and could not replace duties performed by members. The WGA's contract, while more detailed, is essentially similar and states that "AI can't write or rewrite literary material, and AI-generated material will not be considered source material." It also demanded that studios "must disclose to the writer if any materials given to the writer have been generated by AI or incorporate AI-generated material." Their contract also added that the union "reserves the right to assert that exploitation of writers' material to train AI is prohibited."

---

[116] Robert Moran, *Hot Strike Summer*, Brunswick Review, https://review.brunswickgroup.com/article/labor-strike/ (last visited Aug. 22, 2024).
[117] See e.g., Myles Klee & Krystie Lee Yandoli, *Why Striking Hollywood Writers Fear an AI Future*, Rolling Stone, https://www.rollingstone.com/tv-movies/tv-movie-features/wga-strike-hollywood-writers-ai-artificial-intelligence-chatgpt-screenplay-technology-1234728014/ (last visited Aug. 22, 2024).

Voice AI and other generative AI tools are already widely introduced into the Hollywood industries. affected by advances in deep fake technology. Earlier this year, after months of strike action, SAG-AFTRA, the American actor's union, reached an agreement with Replica Studios that would allow voice-over artists to leverage AI technology to monetize the use of their digital voice replicas by studios in video games and other interactive media. The critical value of this agreement is that it allows performers to opt out of the use of their voice in new works. Similarly, the agreement reached between SAG-AFTRA and the Alliance of Motion Picture and Television Producers regarding digitally generated likenesses includes robust provisions relating to consent and compensation.

4. A Path Forward? Principles for a New Social Contract with Technology

The limitations in the current legal framework outlined above, which we would tentatively suggest are typical of most jurisdictions, indicate a need for a new approach. A potential starting point in such a task is to consider a new social contract with technology underpinned by fundamental principles that can guide the development of an appropriate regulatory framework to address the unique challenges posed by voice AI.

Looking forward, it seems inevitable that screen-based interactions with digital technologies will, at some point, be supplanted by voice-based ones. Equally evident is that the existing legal framework lacks adequate mechanisms for regulating this new dominant mode of interaction. The foundation of any new regulatory approach should, therefore, be respect for individual identity and a genuinely human-centric, rather than corporate-centric, perspective. This involves acknowledging the centrality of dominion over one's voice and the principles of respect for individual dignity, autonomy, ownership and control, informed consent, transparency, and accountability and redress.

4.1. Respect for Individual Dignity

The cornerstone of any new social contract with voice AI technologies must be respect for individual dignity. In the context of digital transformation, an individual's voice must be recognized as a fundamental component of their personal identity. As such, shifting from an

enterprise-centric to a human-centric approach is essential.[118] Without this shift, public frustration and mistrust in new technologies will continue to grow, and the enormous potential benefits and opportunities these technologies offer will be overshadowed by the wave of skepticism surrounding technology and, especially, the companies that produce and control these technologies, including Open AI.

Before turning to specific legal principles and rules, it might be important to consider whether and in what ways our concepts of human dignity might require reevaluation. In a post-digital transformation world our identities are more dispersed than ever before. Action-at-a-distance becomes increasingly normalized either on the internet or in virtual, augmented, and other artificial spaces. Consequently, the question arises whether and in what ways our meta-legal concepts like dignity still effectively capture the essence of what it is to be human. Recognition of the profound social, cultural, and economic changes triggered by digital technologies requires us to recognize the limits of extant legal concepts and explore their applicability and meaning in new contexts. The notion of legal certainty, for example, has already been replaced by a more flexible understanding of constant adjustment and similar paralleling potential changes in the concept of dignity.[119]

This more contemporary version of an individual-centric perspective, seen through the lens of a new social contract with technology, would allow us to identify the limits of existing laws and explore the contours of new, desirable regulations appropriate to the realities of our post-DX world.

The transformation of nouns into verbs, such as "Google," for instance, is evident in the realm of generative AI. Voice, once a static asset or characteristic, has become dynamic and activatable for all. The significance of voice and speech in participating in democratic processes and in exercising freedom of thought and expression becomes paramount.

4.2. Ownership and Dominion over One's Voice Data

---

[118] Paulius Jurcys et al., *Data Portability Revisited: Toward the Human-Centric, AI-Driven Data Ecosystems of Tomorrow*, Vanderbilt J. Ent. & Tech. L. (forthcoming Fall 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4475106 (last visited Aug. 22, 2024); Paulius Jurcys, *The Personal AI Revolution: A Human-Centric Approach*, Medium, https://medium.com/prifina/the-personal-ai-revolution-a-human-centric-approach-840f47e92a3a (last visited Aug. 22, 2024).

[119] M. Fenwick, M. Siems & S. Wrbka eds., *The Shifting Meaning of Legal Certainty in Comparative and Transnational Law* (Hart 2020).

In these pivotal times, as technology challenges the old social contract, concepts like autonomy, identity, and ownership become crucial. As evidenced by the various stories in Section 1, voice is profoundly personal. It is integral to how individuals are recognized and distinguished in society, i.e., personal expression through voice is essential to a person's freedom and actualization as a uniquely autonomous subject. Therefore, ownership, dominion, and control over one's voice should be regarded as essential for any individual, particularly in maintaining respect for personal dignity in the context of a post-digital transformation world.[120]

Building on this idea that voice is crucial to personal freedom and our self-realization as individuals, we advocate for the core principle of digital self-determination and dominion over one's own data that powers the creation of derivative voice AI applications. The idea of owning one's voice data is gaining prominence.[121] Even though in the recent legal literature and regulatory policies, the notion of individual ownership of personal data has been criticized and mostly rejected; we would urge a reconsideration of this issue.[122]

The inadequacy of the existing framework necessitates the establishment of new rules that respect identity and autonomy. In the digital era, data and voice are inseparable from digital autonomy. In the new social contract and data economy, individuals own their data and, thus, their voice. Instead, we would argue that individual ownership over voice data is a necessary precondition to establishing more specific rules about utilizing such data by giving access and control to third parties that could integrate such data in various AI-powered applications and environments.

Establishing clear rules over individual data ownership enables new forms of consent, data access, opt-in, and opt-out frameworks.[123] As digital UX/UI interfaces increasingly rely on voice commands, it becomes sensible to design new frameworks and rules to prevent the misuse of someone else's voice. The concept of ownership is fundamental in society, forming the basis for contract and tort law, and enabling the licensing or reclamation of rights. Ownership of one's voice is a foundational building block, providing the legal certainty needed to preserve autonomy and dignity.

---

[120] Paulius Jurcys et al., *Ownership of User-Held Data: Why Property Law Is the Right Approach*, Harv. J.L. & Tech. Dig. (Sept. 21, 2021),
https://jolt.law.harvard.edu/digest/ownership-of-user-held-data-why-property-law-is-the-right-approach.

[121] Paulius Jurcys et al., *Ownership of User-Held Data: Why Property Law Is the Right Approach*, Harv. J.L. & Tech. Dig. (Sept. 21, 2021),
https://jolt.law.harvard.edu/digest/ownership-of-user-held-data-why-property-law-is-the-right-approach.

[122] Paulius Jurcys et al., *Ownership of User-Held Data: Why Property Law Is the Right Approach*, Harv. J.L. & Tech. Dig. (Sept. 21, 2021),
https://jolt.law.harvard.edu/digest/ownership-of-user-held-data-why-property-law-is-the-right-approach.

[123] Paulius Jurcys et al., *My Data, My Terms: A Proposal for Personal Data Use Licenses*, Harv. J.L. & Tech. Dig. (Mar. 5, 2020), https://jolt.law.harvard.edu/digest/my-data-my-terms.

4.3. Consent as a Default Rule

From a legal perspective, the default principle to be derived from the above considerations is that an individual's express consent is a necessary precondition for the deployment of Voice AI technologies. This principle should apply universally, ensuring that individuals must opt-in to voice cloning frameworks with full awareness of the implications. Emerging industry standards, especially for direct-to-consumer products, already adopt a similar approach based on users consent to cloning their voices and live with the consequences.[124]

Yet, current laws, particularly those focused on consumer and market protection, do not sufficiently address the issue of voice protection. The previous section demonstrated these limitations, and such a critique of existing law underscores the need for a new social contract to address these emerging social relations. In thinking about new rules that specify the implementation of consent in various practical scenarios, it is important that policy makers, regulators, individuals, and other stakeholders collaborate closely with voice AI technology developers. Such collaboration is essential to align the interests of various stakeholders and their incentives with the emerging business models that rely on the use of voice AI technologies.

4.4. Transparency & Disclosure

Transparency and disclosures are also critical, especially given the technological, behavioral, and legal shortcomings observed in personal data regulations. For non-B2C products, individual opt-in may not always be feasible in commercial settings. More generally, bad actors will inevitably find ways to create deep fakes without an individual's consent. One possible way to address this issue of informed consent could be a requirement of transparency with regard to the use of data. What data has been used as an "input" to train voice AI models? How do voice AI models operate? What data has been used to generate voice AI outputs?

Transparency requirements and disclosures could be a powerful tool to ensure that voice AI technologies are developed and deployed responsibly. Voice AI developers should ensure that the terms of use of their products incorporate the requirement of explicit consent for voice replication or modification and include tools to clearly indicate AI-generated output. The principle that individuals own their voice is emerging, as evidenced by leading companies like Eleven Labs. From the individual perspective, the principle of digital self-determination should be at the core of this new legal framework. Individuals must have the right to control how their voice is used in digital environments. This includes the right to consent to voice replication, the

---

[124] See terms of use of Eleven Labs and LMNT.

right to be informed about how their voice data is used, and the right to withdraw consent at any time.

4.5. Accountability and Redress

The new legal framework should also include mechanisms for accountability and redress. If their voice rights are violated, individuals should be able to seek redress. This could include legal action against those who misuse voice technologies and mechanisms to compensate individuals for any harm suffered. To implement these principles, we need clear, enforceable rules. These rules should be developed through a participatory process that includes input from a wide range of stakeholders, including technologists, legal experts, policymakers, and the public. This participatory approach will ensure that the rules are balanced, fair, and reflective of societal values.

Existing law is reactive and thus focused on compensation and retribution, whereas emerging technologies provide opportunities to create new uses for voice that are more prospective-transactional. Voice undoubtedly falls within the scope of the law, but we have not yet envisioned new forms of regulation appropriate to this issue.

Impersonation in the age of AI raises novel and difficult questions about whether we should equate impersonation and cloning. Impersonation is typically more akin to caricature, something that falls within the scope of freedom of expression, while cloning is akin to unlawful copying. Moreover, in an era of generative AI, voice cloning has a more permanent dimension, allowing for continuous reuse and reproduction. The old legal frameworks were designed for impersonation cases, but in the context of generative AI, a new social contract is needed. This quantitative difference in technology use may lead to qualitative differences in regulatory needs.

One practical illustration of a consent-based approach to synthetic voice coupled with practical remedies could be found in the recent US Copyright Office consultation on Copyright and Generative AI.[125] Since the proliferation of voice cloning technologies, voice actors, for instance, have started entering into licensing agreements to protect their voices in the age of generative AI. Various stakeholders have identified the need for federal legislation to address AI impersonation, such as the proposed Federal Anti-Impersonation Right Act (FAIR Act), recently proposed by Adobe.[126] This act would grant artists the right to take legal action against those who

---

[125] *Artificial Intelligence Study*, U.S. Copyright Office, https://www.copyright.gov/policy/artificial-intelligence/ (last visited Aug. 22, 2024).
[126] *Comment from Adobe Inc.*, Comment ID COLC-2023-0006-8594, https://www.regulations.gov/comment/COLC-2023-0006-8594 (last visited Aug. 22, 2024).

intentionally and commercially impersonate their work and style using AI tools, providing a new mechanism to protect their livelihood without relying solely on copyright and fair use laws. The FAIR Act aims to prevent a patchwork of conflicting state laws and includes provisions for statutory damages to reduce the burden on plaintiffs.

5. Conclusion

The interplay between emerging technologies and law has always been dynamic. However, the pace at which generative AI and voice technologies are evolving necessitates a more proactive legal approach. Traditional legal principles must adapt to contemporary realities. The role of voice in identity, privacy, and autonomy is becoming increasingly significant in legal discourse. The importance of voice extends beyond individual recognition to encompass broader societal implications. Voice technologies have the potential to enhance or undermine democratic participation and freedom of expression. Therefore, the legal framework must ensure that these technologies are used in ways that uphold democratic values and human rights.

In this paper we illustrated that current laws are inadequate to address the issues arising in the age of generative AI technologies, which introduce new questions. Generative AI and its implications for voice cloning illustrate the pressing need for a new social contract. While existing legal frameworks offer some guiding principles regarding impersonation, they fail to recognize that voice, as a vital component of personal identity, should perhaps fall within general tort law. However, tort law, while providing a backstop for wrongs that fall through the net of other legal fields, is often not very practical in addressing new technological uses of voice.

In light of the rapid advancement of voice AI technologies, the need for a new social contract and legal framework becomes obvious. This framework should be grounded in respect for individual dignity, autonomy, and digital self-determination. By developing clear, enforceable rules and mechanisms for accountability and redress, we can ensure that voice technologies are used in ways that benefit individuals and society while mitigating the risks of misuse. Through education, awareness, and international cooperation, we can build a future where voice technologies enhance human potential and uphold democratic values.

This new legal framework must be dynamic and adaptable, capable of evolving with technological advancements. It should anticipate future developments and be ready to address new challenges as they arise. By fostering a proactive and forward-thinking approach, we can ensure that the legal system remains relevant and effective in the face of rapid technological change. Lastly, the concept of digital self-determination must be expanded to include not just control over one's voice but also control over all forms of personal data.

To facilitate the development of such a new social contract, two additional practical strategies could be identified and pursued: First, it is important to facilitate education and AI literacy within the global community. As voice technologies become more prevalent, it is important to teach every individual how to integrate AI tools into daily lives and professional environments. Improving AI literacy also includes understanding the risks associated with voice technologies and facilitating the realization of the need to have informed consent-based interactions with voice AI technologies. Second, international cooperation is essential. Voice technologies are global in nature, and their regulation requires coordinated efforts across jurisdictions. International standards, bottom-up consensus among industry players, and global guidelines can help ensure that voice rights are equally respected and protected consistently worldwide.