

**IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE**

| | | |
|--|---|-----------------------------|
| KARISSA VACKER, MARK BOYETT, |) | |
| BRIAN LARSON, IRON TOWER PRESS, |) | |
| INC., and VAUGHN HEPPNER, |) | |
| |) | C.A. |
| Plaintiffs, |) | |
| |) | |
| v. |) | JURY TRIAL REQUESTED |
| |) | |
| ELEVEN LABS, INC., |) | |
| |) | |
| Defendant. |) | |

COMPLAINT

Plaintiffs KARISSA VACKER, MARK BOYETT, BRIAN LARSON, IRON TOWER PRESS, INC. AND VAUGHN HEPPNER (collectively, “Plaintiffs”) file this Complaint for misappropriation of their likenesses and publicity rights and violations of the Digital Millennium Copyright Act against Defendant, ELEVEN LABS, INC. (“ElevenLabs” or “Defendant”).

INTRODUCTION

Plaintiffs Ms. Karissa Vacker (“Vacker”) and Mr. Mark Boyett (“Boyett”) (collectively, the “Voice-Actor Plaintiffs”) are highly accomplished professional voice actors. Plaintiffs Brian Larson (“Larson”), Iron Tower Press, Inc. (“ITP”) and Vaughn Heppner (“Heppner”) (collectively the “Author/Publisher Plaintiffs”) are authors and publishers that own copyrights in works narrated by Mr. Boyett. Defendant ElevenLabs offers text-to-speech services through its website and Application Programming Interface (“API”), allowing users to create synthetic audio narrations of written text. Defendant’s system employs generative artificial intelligence to clone voices, synthesize speech, and perform multilingual voice-to-voice translation.

As professional voice actors, Vacker and Boyett rely on their distinctive voices and speaking styles as the foundation of their livelihoods and professional identities. ElevenLabs

misappropriated the Voice-Actor Plaintiffs’ voices and likenesses by, among other things, creating voice clones that it named “Bella” and “Adam,” respectively, and is using them without Voice-Actor Plaintiffs’ consent or compensation to promote its services, attract millions of users, and generate significant revenue, market value, and investment. These voice clones mimic Plaintiffs’ distinctive vocal timbres, accents, intonation, pacing, vocal mannerisms, and speaking styles—delivering a synthetic professional narration that friends and family recognize as their voices, allowing ElevenLabs to profit from Voice-Actor Plaintiffs’ valuable assets.

In the process of creating Bella and Adam, ElevenLabs also violated the Digital Millennium Copyright Act (“DMCA”) by circumventing technological measures protecting Author/Publisher Plaintiffs’ copyrighted audiobook narrations and removing or altering copyright management information to train its AI system without authorization.

ElevenLabs is liable to Plaintiffs for misappropriation of likeness, identity, and publicity rights under Texas and New York law, as well as violations of the DMCA’s anticircumvention provisions and Copyright Management Information (17 U.S.C. §§ 1201 and 1202); Plaintiffs seek injunctive relief, damages, and other remedies.

PARTIES

1. Plaintiff Karissa Vacker is an award-winning professional voice actress who creates audiobook narrations of published written works for publishers who distribute them on common audiobook subscription services and through digital retail outlets. In addition to audiobooks, Vacker provides vocal talent for audio narrations used in filmic works and advertisements. She has produced a voluminous body of work that is commercially available over the internet. Vacker is a citizen and resident of Texas.

2. Plaintiff Mark Boyett is a voice actor who creates audiobook narrations of published books for a number of publishers who distribute them on common audiobook

subscription services and other digital retail outlets. In addition, Boyett's vocal talent is also used in filmic works and advertisements. Boyett is a citizen and resident of New York.

3. Plaintiff Brian (BV) Larson is an American author renowned for his extensive science fiction catalog. With a prolific output comprising more than 70 books and 4 million copies sold, Larson has established himself as a USA Today bestselling author. His work spans several series, including "Haven," "Imperium," "Star Force," "Hyborean Dragons," and "Unspeakable Things," showcasing his versatility across different sub-genres. Most of his titles have been translated into other languages, further broadening his international readership. Most of his titles have been adapted into audiobooks, including over 40 audiobooks narrated by Boyett. Larson owns the copyright in the underlying written work for all audiobook narrations of his work (and, for some but not all of these works, he also owns the copyright in the audiobook recording) narrated by Boyett. Larson's written work and audio narrations by Boyett are distributed through online digital retailers and audiobook subscription services. Larson is a citizen and resident of California.

4. Plaintiff Iron Tower Press, Inc. is a California corporation owned and operated by Larson that publishes written works and commissions audiobook narrations. ITP is a corporation organized and based in California.

5. Plaintiff Vaughn Heppner is a prolific author residing in Nevada who primarily writes in the military science fiction, space opera, and thriller genres. His works frequently appear on bestseller lists for these genres on internet retailers like Amazon. He has published over 10 major novel series, including his Lost Starship series. Boyett has narrated 20 of the 21 books in Heppner's Lost Starship series for audiobooks. Heppner owns the copyrights in the underlying written work and for all audiobook narrations performed by Boyett.

6. Defendant ElevenLabs, Inc. is a corporation organized and existing under the laws

of the State of Delaware, with a primary mailing address at 576 Vanderbilt Avenue, Apt. 4, New York City, New York 11238, USA. On information and belief, its principal place of business is in New York City, New York. ElevenLabs offers text-to-speech services, including the capability to clone the user's voice and render text to speech in that voice. Defendant provides several default high quality voices in which it can render text to speech. Defendant does business throughout the United States, including in this District.

JURISDICTION

7. This Court has subject matter jurisdiction over this action pursuant to 28 U.S.C. §§ 1331 and 1338(a), as it arises under the DMCA of 1976, 17 U.S.C. § 101, et seq.

8. This Court has personal jurisdiction over ElevenLabs because ElevenLabs is a Delaware corporation.

VENUE

9. Venue in this Court is proper under 28 U.S.C. § 1391(b) because ElevenLabs is a Delaware corporation.

FACTS

A. The Voice-Actor Plaintiffs Are Professional Voice Actors Who Specialize in Creating Audiobook Narrations.

1. Karissa Vacker

10. Karissa Vacker is an award-winning voice actress who uses her distinctive voice to earn her livelihood. She has produced a voluminous body of work, including narrations of more than 300 audiobooks for major publishers including Penguin Random House, MacMillan Audio, Hachette, Brilliance, Tantor, Dreamscape, HarperCollins, and Scholastic. Vacker produces work under her own name, Karissa Vacker, and under her artist name, Vanessa Edwin.

11. Hundreds of audiobook narrations performed by Vacker are commercially

available for purchase and subscription-based access online through multiple vendors, including over 200 audiobook narrations available on Amazon's Audible service alone.

12. Vacker's voice is distinctive, recognizable, and expressive, and has earned her acclaim and recognition in the audiobook industry.

13. Vacker's distinctive voice and vocal style is a valuable asset and source of income, which Vacker earns by charging fees for her services narrating text in her distinctive voice and vocal style.

14. Vacker has also narrated text for audio-visual productions, including fiction and non-fiction works, as well as advertisements.

15. Vacker has also worked as an on-screen actress, primarily in major network television shows, including "How I Met Your Mother," "Chuck," "Castle," "Days of Our Lives," and "NCIS." Vacker has also played roles in films, including "Someone Marry Barry" and "Summer Forever," and is featured in a popular video game, "Fight Night Champion," where she played the role of a reporter.

16. Vacker has not licensed, authorized, or consented to ElevenLabs digitally "cloning" her voice through its text-to-speech system, nor has she authorized or consented to ElevenLabs distributing audio recordings of text narrated in her distinctive voice and likeness.

17. Vacker has not authorized ElevenLabs to use, copy, or reproduce original recordings of her vocal performances, generate new audio recordings that use her distinctive voice and likeness, or otherwise use her voice and likeness for any purpose.

2. Mark Boyett

18. Mark Boyett is an award-winning voice actor who has been creating audiobook narrations of published books for many years. Boyett has narrated an extensive catalog of works spanning a range of literary genres.

19. Boyett's distinctive and recognizable voice and speaking style has earned him acclaim and recognition in the audiobook industry and is a valuable asset and source of income, which Boyett earns by charging fees for his narration services.

20. Boyett has produced a voluminous body of audiobook narrations available for purchase and subscription-based access online through multiple vendors, including Amazon's Audible service. Boyett has also voiced the introduction to audiobooks produced by Amazon's Audible Studios and Audible Originals since 2018.

21. Boyett also has worked as an on-screen film actor, appearing in 17 films and TV shows, including appearances in the television series "Blue Bloods," "The Gilded Age," "FBI," "Law & Order: Special Victims Unit," and others. He has appeared in feature films including "The Tender Bar," and "Fell, Jumped or Pushed."

22. Mr. Boyett has not licensed, authorized, or consented to ElevenLabs digitally "cloning" his voice through its text-to-speech system, nor has he authorized or consented to ElevenLabs distributing audio recordings that narrate text in his distinctive voice and likeness.

B. The Author/Publisher Plaintiffs are copyright owners in audiobook recordings and/or written works narrated by the Voice-Actor Plaintiffs.

23. Plaintiffs Brian (BV) Larson, Iron Tower Press, Inc., and Vaughn Heppner are authors and copyright owners of certain audiobook recordings performed by Boyett and/or the underlying written work narrated by him, distributed through online retailers and subscription services like Amazon's Audible.

1. Brian (BV) Larson.

24. Plaintiff Brian (BV) Larson is an American author renowned for his extensive work in science fiction. With a prolific output of more than 70 books and over 4 million copies sold, Larson has established himself as a USA Today bestselling author. His work spans several series,

including “Haven,” “Imperium,” “Star Force,” “Hyborean Dragons,” and “Unspeakable Things,” showcasing his versatility across different sub-genres within science fiction and fantasy. Most of his titles have been translated into other languages, further broadening his international readership.

25. Most of Larson’s books have been adapted into audiobooks and published in audiobook format, which are distributed through online digital retailers and audiobook subscription services.

26. Boyett, has performed audiobook narrations of numerous works authored by Larson that are distributed through online digital retailers and audiobook subscription services. Boyett has narrated over 40 works authored by Larson on Amazon's Audible service alone, including: *Steel World; Dust World; Home World; Tech World; Swarm; Rogue World; Machine World; Death World; Blood World; Dark World; Rebel Fleet; Armor World; Extinction; Conquest; Rebellion; Battle Station; Storm World; Orion Fleet; Empire; Starship Liberator; Annihilation; Alpha Fleet; Storm Assault; Ice World; The Dead Sun; Outcast* (coauthored with David VanDyke), *Battleship Indomitable* (coauthored with David VanDyke), *Exile* (coauthored with David VanDyke), *Earth Fleet; City World; Demon Star* (coauthored with David VanDyke), *Flagship Victory* (coauthored with David VanDyke), *Hive War* (coauthored with David VanDyke), *Straker’s Breakers* (coauthored with David VanDyke), *Starship Pandora; Army of One; War of the Spheres* (coauthored with James D. Millington), *Amber Magic; Planetary Assault* (coauthored with Vaughn Heppner and David VanDyke), *Dragon Magic; Death Magic; Sky Magic; Blood Magic; Dream Magic; and Shadow Magic.*

27. Larson owns the copyrights in the underlying text of all audiobook narrations of his work narrated by Boyett, and most of the audiobook narrations as well.

28. Many of Larson's written works have been translated into other languages, including German and Polish, and many of those have been recorded and sold as audiobooks.

2. Vaughn Heppner

29. Vaughn Heppner is a prolific author who primarily writes in the military science fiction, space opera, and thriller genres. His works frequently appear on bestseller lists for these genres on internet retailers like Amazon. His works span over 10 novel series, the most popular of which is his Lost Starship series containing 20 novels. Boyett has narrated the audiobooks for 19 of the 20 works in the Lost Starship series, including: *The Lost Command*, *The Lost Destroyer*, *The Lost Colony*, *The Lost Patrol*, *The Lost Planet*, *The Lost Earth*, *The Lost Artifact*, *The Lost Star Gate*, *The Lost Supernova*, *The Lost Swarm*, *The Lost Intelligence*, *The Lost Tech*, *The Lost Secret*, *The Lost Barrier*, *The Lost Nebula*, *The Lost Relic*, *The Lost Task Force*, *The Lost Clone*, *The Lost Portal*, and *The Lost Cyborg*.

30. Heppner owns the copyrights in both the underlying written work and audiobook narrations for all works authored by him and narrated by Boyett. These works are distributed through online retailers, including Amazon's Audible service.

C. ElevenLabs is a generative AI company that launched its platform offering text-to-speech services in January 2023 and is now valued at over \$1 billion.

1. ElevenLabs attracted over a million new users in its first six months by giving away free audio narrations narrated by its "default" voices, including "Adam" and "Bella."

31. ElevenLabs is a software company that specializes in developing natural-sounding speech synthesis and text-to-speech software using AI and deep learning. ElevenLabs provides text-to-speech services to users through its website, elevenlabs.io, and through its API. Its service allows users to create synthetic audio narrations of a written text supplied by the user.

32. ElevenLabs was co-founded in 2022 by Piotr Dąbkowski, an ex-Google machine

learning engineer, and Mateusz Staniszewski, an ex-Palantir deployment strategist.

33. As ElevenLabs explained in an early blog post, its goal was to create a text-to-speech system that could generate “*voice actor-grade speech* from any text”:¹

Who’s it for?

We chose this direction for a number of reasons. There is currently no tool which supports generating long-form speech in high enough quality to make it suitable for voicing news or audiobooks. Our team are keen listeners of all things audio and we felt that rising to the challenges posed by lengthier content is a natural step towards realising our ambitions. But we’re also particularly excited to consider it our stand-out feature - we’re the first AI speech tech platform to bring the most emotive, rich and lifelike voices to creators and publishers seeking the ultimate storytelling quality.

To this extent, our platform allows you to generate and download high quality, *voice actor-grade speech* from any text - be it news articles, books, newsletters, blogs or academic papers.²

34. ElevenLabs raised \$2 million from investors in a pre-seed round, and in January 2023 publicly released its beta platform.³

35. Upon release, ElevenLabs offered nine default voices, including Adam and Bella.⁴

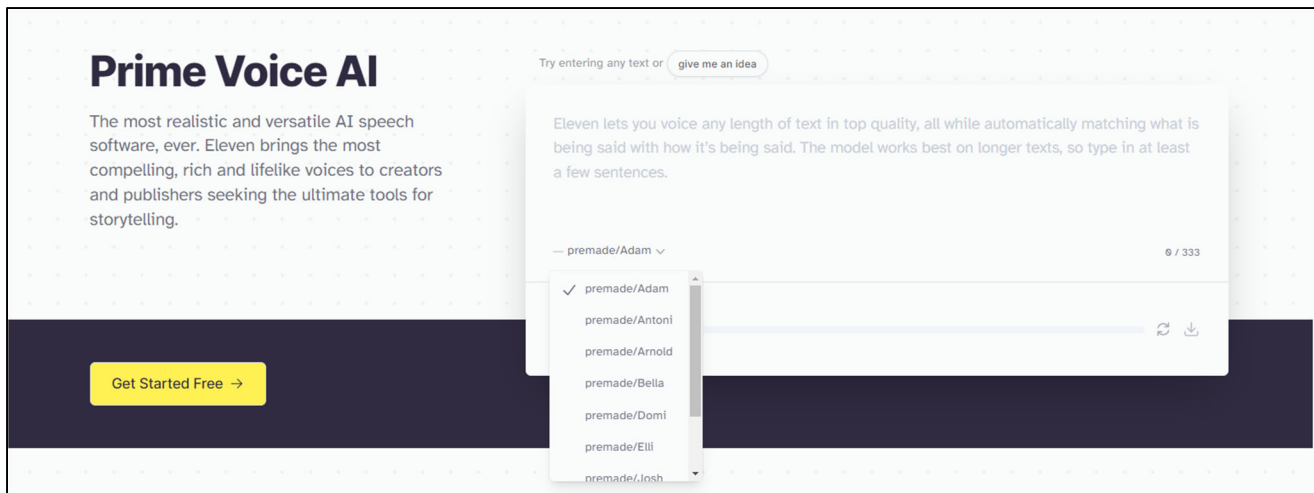
36. The landing page for the ElevenLabs website invited visitors to try its services by entering text and obtaining a high-quality professional audio narration of the text rendered in one of its nine default voices, at no cost:

¹ ElevenLabs Website, Blog, October 17, 2022 post by ElevenLabs Team (available at <https://elevenlabs.io/blog/long-form-speech-synthesis-for-publishers-and-creators/>).

² *Id.* (emphasis added).

³ Wikipedia article on ElevenLabs, *available at* <https://en.wikipedia.org/wiki/ElevenLabs> (last accessed February 7, 2024).

⁴ *See, e.g.*, Internet Archive WaybackMachine, *Elevenlabs.io* Landing Page, archived March 5, 2023, *available at* <https://web.archive.org/web/20230305062351/https://beta.elevenlabs.io/>.



37. By June 2023, ElevenLabs had attracted over 1 million registered users.⁵

38. That same month, ElevenLabs raised \$19 million in a Series A funding round at a company valuation of roughly \$100 million, despite the company having no office and only 15 employees.⁶

2. By January 2024, ElevenLabs had attracted millions of users—resulting in a company valuation of \$1.1 billion.

39. By January 2024, ElevenLabs’ service was being used by “millions,”⁷ allowing it to raise an additional \$80 million in series B funding based on a market valuation of \$1.1 billion leading to self-proclaimed “unicorn status.”⁸

⁵ ElevenLabs Website, Blog Entry dated June 20, 2023 (“AI platform ElevenLabs, which launched in Beta in January 2023, has now attracted over 1 million registered users across creative, entertainment and publishing spaces”), available at <https://elevenlabs.io/blog/elevenlabs-launches-new-generative-voice-ai-products-and-announces-19m-series-a-round-led-by-nat-friedman-daniel-gross-and-andreessen-horowitz/> (last accessed March 14, 2024).

⁶ *Id.*; see also Wikipedia, article on ElevenLabs, available at <https://en.wikipedia.org/wiki/ElevenLabs> (last accessed February 7, 2024) (“In June 2023, ElevenLabs raised a \$19 million Series A funding round at a valuation of about \$100 million,[7][8] despite the company having no office and only 15 employees.”); ElevenLabs Website, Blog Entry, Jan. 22, 2024 (“ElevenLabs has raised an \$80m Series B round co-led by Andreessen Horowitz, Nat Friedman, Daniel Gross, and joined by Sequoia Capital, Smash Capital, SV Angel, BroadLight Capital and Credo Ventures to strengthen its position as the leader in voice AI.... Since its launch, ElevenLabs technology has [been] adopt[ed] by millions leading the company to unicorn status.”), available at <https://elevenlabs.io/blog/series-b/> (last accessed March 4, 2024).

⁷ ElevenLabs Website, Blog Entry, Jan. 22, 2024, available at <https://elevenlabs.io/blog/series-b/> (last accessed March 4, 2024).

⁸ *Id.*

40. According to a January 2024 blog post on the ElevenLabs website, the company had generated over **100 years of audio** since its launch, and its technology was used by employees at 41% of Fortune 500 companies:⁹

Since its public launch, ElevenLabs has led the industry in natural speech synthesis, enabling users to create and design AI voices across a vast swathe of languages and accents, with the ability to deliver a wide range of emotions and intonations. Since launch, ElevenLabs' users have generated over **100 years** of audio, while the firm grew from 5 to **40 employees**. Today, ElevenLabs technology is being used by employees at **41% of Fortune 500** companies.

3. ElevenLabs' Services

(a) ElevenLabs' "Speech Synthesis," first in English and then Multi-Lingual

41. ElevenLabs is known for its browser-based, "Speech Synthesis" AI-assisted text-to-speech software, which can produce lifelike speech by synthesizing vocal emotion and intonation, using its default voices, including "Bella" and "Adam."¹⁰ The system also includes the ability to "clone" the voice of the user from a sample provided by the user (*i.e.*, to identify the unique set of features of the voice in the audio sample provided by the user by mapping them to features of voices learned during training from the training data).

42. Originally, the system could only convert text to speech in English. On April 27, 2023, ElevenLabs released the "Eleven multilingual V1," supporting text-to-speech in "seven new languages: French, German, Hindi, Italian, Polish, Portuguese, and Spanish."¹¹

43. In August 2023, ElevenLabs released its Multilingual V2 model, supporting text-

⁹ *Id.*

¹⁰ Wikipedia article on ElevenLabs, *available at* <https://en.wikipedia.org/wiki/ElevenLabs> (last accessed February 7, 2024).

¹¹ See ElevenLabs Website, Blog Entry, April 27, 2023, Introducing Eleven Multilingual v1: Our New Speech Synthesis Model (*available at* <https://elevenlabs.io/blog/eleven-multilingual-v1>

to-speech in **28 languages**, adding Chinese, Korean, Dutch, Turkish, Swedish, Indonesian, Filipino, Japanese, Ukrainian, Greek, Czech, Finnish, Romanian, Danish, Bulgarian, Maylay, Slovak, Croatian, Classic Arabic, and Tamil.¹² With the multilingual model, the user can narrate text using any of ElevenLabs' default voices (or use a voice cloned from a short audio sample provided by the user), emulating the voice, prosody, pacing, and style, in any of the supported languages as if the person was perfectly bilingual.

(b) Dubbing Studio (Speech to Speech Translation)

44. On January 22, 2024, ElevenLabs announced its Dubbing Studio product, “which enables users to dub entire movies, as well as generate and edit their transcripts, translations, and timecodes, providing additional control over content production. These capabilities supplement the already existing AI dubbing feature which enables automated, end-to-end video localization across 29 languages.”¹³ The dubbing tool enables the user to automatically translate (audio) speech in one language to speech in the same voice in another language, while preserving the original voice, prosody, pacing, style, rhythm and timing of the delivery, while preserving the emotional character of the performance.¹⁴

45. ElevenLabs has achieved substantial and rapid growth, wealth, value, and profits from the creation, broadcast/performance, and distribution of synthetic narrations created using its “Bella” and “Adam” voices. Moreover, ElevenLabs granted its paying subscribers a *worldwide license* to use, perform, and redistribute these recordings *royalty-free, for commercial purposes*.

¹² See ElevenLabs Website, Blog Entry, August 22, 2023, by ElevenLabs Team Multilingual v2, (*available at <https://elevenlabs.io/blog/multilingualv2/>*).

¹³ *Id.*

¹⁴ ElevenLabs Website, Blog Entry, Jan. 22, 2024, available at <https://elevenlabs.io/blog/series-b/> (last accessed February 7, 2024); Wikipedia, article on ElevenLabs, *available at <https://en.wikipedia.org/wiki/ElevenLabs>* (last accessed February 7, 2024).

4. During this period, the U.S. government issued a call to action to help mitigate the harm caused by new state-of-the-art text-to-speech systems.

46. Unfortunately, such technology also enables users to misappropriate the likeness of others, with troubling consequences.¹⁵

47. For example, in January 2024, a “deepfake” audio recording in the voice of US President Joe Biden was transmitted to roughly 5,000 to 25,000 New Hampshire primary voters via “robocalls,” urging them not to vote in the New Hampshire primary election.¹⁶ Only after an *independent* investigation traced the recording to ElevenLabs, the company suspended the user’s account.¹⁷

48. According to *Fortune Magazine*, ElevenLabs was allowing users to create voice clones of public figures, including politicians, with the understanding that the audio must “express humor or mockery in a way that it is clear to the listener that what they are hearing is a parody.”¹⁸

49. In addition to ElevenLabs’ failure to implement adequate safeguards to police against foreseeable abuse by its *users*, ElevenLabs *itself* has misused its own technology, from

¹⁵ See, *Forbes*, September 3, 2019, “A Voice Deepfake Was Used To Scam A CEO Out Of \$243,000,” available at <https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=7d58ec102241>.

¹⁶ *Fortune Magazine*, Gargi Murphy, et. al., *Biden audio deepfake spurs AI startup ElevenLabs—valued at \$1.1 billion—to ban account: ‘We’re going to see a lot more of this’* (January 27, 2024), available at, <https://fortune.com/2024/01/27/ai-firm-elevenlabs-bans-account-for-biden-audio-deepfake/>; see also Associated Press, *Magician says political consultant hired him to create AI robocall ahead of New Hampshire primary*, Holly Ramer et. al, (February 23, 2024), available at <https://apnews.com/article/biden-robocalls-ai-magician-new-hampshire-louisiana-155b3ffe9d24048f3380104f95b48a57>; Associated Press, Holly Ramer, *Political Consultant behind fake Biden robocalls says he was trying to highlight the need for AI rules* (February 26, 2024) (available at <https://apnews.com/article/ai-robocall-biden-new-hampshire-primary-2024-f94aa2d7f835ccc3cc254a90cd481a99>); February 7, 2024, NPR, All Things Considered, *What a robocall of Biden’s AI-generated voice could mean for the 2024 election*, <https://www.npr.org/2024/02/07/1229856682/what-a-robocall-of-bidens-ai-generated-voice-could-mean-for-the-2024-election> (estimating number robocalls were made to 5,000 – 25,000 homes).

¹⁷ *Fortune Magazine*, Gargi Murphy, et. al., *Biden audio deepfake spurs AI startup ElevenLabs—valued at \$1.1 billion—to ban account: ‘We’re going to see a lot more of this’* (January 27, 2024), available at, <https://fortune.com/2024/01/27/ai-firm-elevenlabs-bans-account-for-biden-audio-deepfake/>.

¹⁸ *Fortune Magazine*, Gargi Murphy, et. al., *Biden audio deepfake spurs AI startup ElevenLabs—valued at \$1.1 billion—to ban account: ‘We’re going to see a lot more of this’* (January 27, 2024) (available at, <https://fortune.com/2024/01/27/ai-firm-elevenlabs-bans-account-for-biden-audio-deepfake/>).

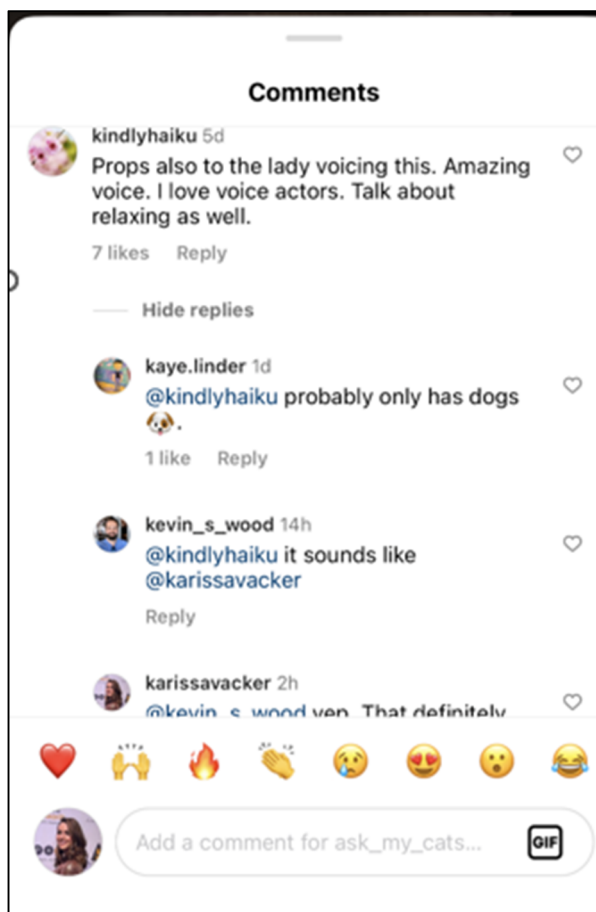
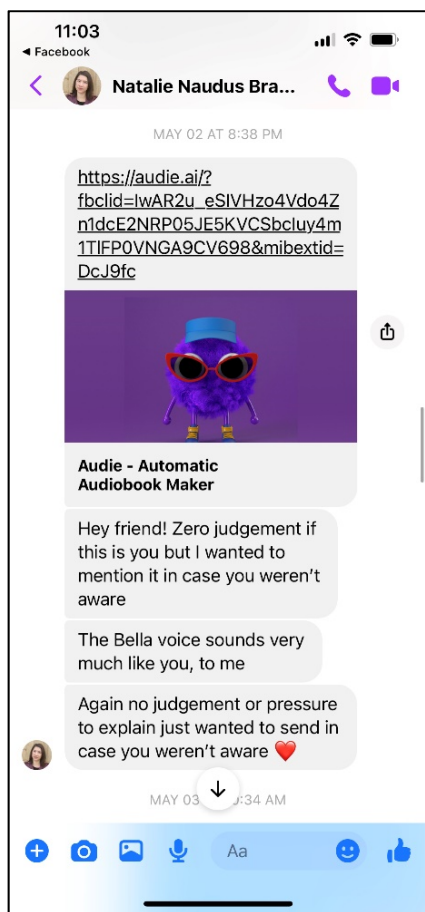
inception, by using copyrighted works without authorization, stripping them of their digital rights protections, and using them to clone the voice and likeness of Vacker and Boyett. ElevenLabs did this without the Voice-Actor Plaintiffs' knowledge or permission and has used their voices to promote its services and attract millions of users by offering professional-quality audio narrations.

50. ElevenLabs sold—and on information and belief, continues to sell—text-to-speech services rendered in the voices and likenesses of Vacker and Boyett.

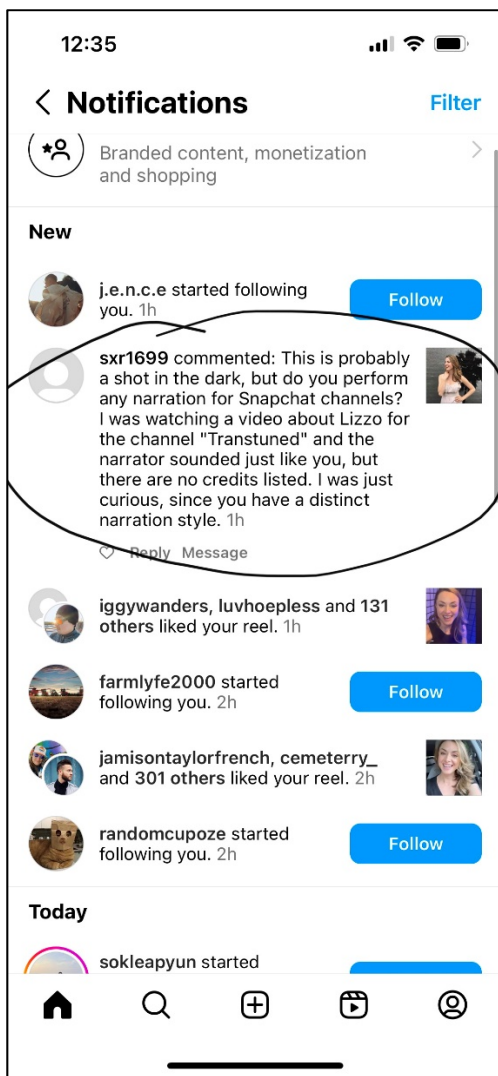
D. ElevenLabs misappropriated the voices and likenesses of the Voice-Actor Plaintiffs.

1. Vacker learned from friends and fans of her work that ElevenLabs had cloned her voice and was distributing recordings using her voice and likeness.

51. In late 2023, Vacker learned from fans that ElevenLabs had digitally cloned her voice and was distributing recordings performed in her unique likeness:



52. Vacker was shocked and distressed to discover that her voice had been cloned without her knowledge or consent.



53. Vacker found that the “Bella” voice offered by ElevenLabs imitated her own voice, capturing her unique vocal qualities, intonation, and style. The similarity was so striking that her friends, family, and colleagues recognized the “Bella” voice as Vacker’s.

54. Vacker reached out to ElevenLabs to demand an explanation for how they had obtained her voice data and to request that they cease using her voice immediately. In mid-November, 2023, ElevenLabs promised to remove the “Bella” voice from its website and prevent

its users from creating new content in Vacker’s voice and likeness.

55. While ElevenLabs removed Bella from its website, Vacker continued to encounter new recordings on YouTube, social media, and elsewhere narrated in her distinctive voice and style.

56. Upon investigation, Vacker discovered that third-party services, including Audie.ai—a website that provided text-to-speech services to its own customers using the ElevenLabs API¹⁹ services—continued to allow their users to select the “Bella” voice, create new content in Vacker’s voice and likeness, and use these recordings in projects over which she has no control, to promote messaging she did not approve, without her approval or any compensation.

57. On information and belief, ElevenLabs continued to allow its customers accessing ElevenLabs through the API to create content using the “Bella” voice.²⁰

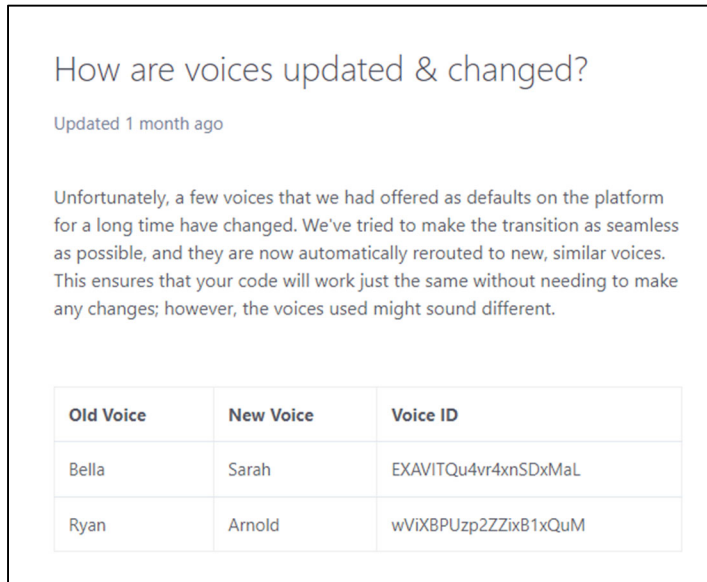
58. In December 2023, Vacker again reached out to ElevenLabs, explaining that she had discovered ElevenLabs was still generating and selling synthetic recordings using its default “Bella” voice to customers via the API. ElevenLabs claimed that removing API access to Bella was a technical challenge that ElevenLabs was not sure it could overcome.

59. On December 28, 2023, ElevenLabs confirmed that Bella had been removed from the ElevenLabs API. On information and belief, ElevenLabs re-assigned the “voice_id” number used to select the “Bella” voice when making an API call (voice_id

¹⁹ Authors Voice Website (accessible at <https://audie.ai/>); *see also id.* at “Thank You” page (available at <https://audie.ai/thank-you/>) (“Audie has been a long time partner of Elevenlabs.io, the originators of the incredible AI voices you can use on Audie. They recently improved their ‘projects’ system so much that it can now read entire audiobooks. With controllable emotion! They can even translate it into different languages, automatically, with AI. Now you can give it a try! The same voices you enjoyed here at Audie are available there, in fact there’s thousands more.”).

²⁰ An API, or Application Programming Interface, is a tool that allows different software programs to talk to each other. It is used to let one program use services or data from another program. For example, if a software program wants to use a service provided by another company, it can use an API to connect to that service and use it within its own application. This connection is made through something called API calls.

“EXAVITQu4vr4xnSDxMaL”) to point to a new and different default voice named “Sarah.”²¹




60. Nonetheless, Vacker continues to encounter recently-created recordings narrated in her own voice on video sites like YouTube, social media, and elsewhere throughout the internet.





61. On information and belief, ElevenLabs’ users, unhappy with the removal of the Bella voice, have used ElevenLabs’ “voice-cloning” capability to create their own high-quality clones of the Bella voice. These voice clones are made available for collective use in ElevenLabs’ VoiceLab library with names like “ReBella,” “Belle,” and “Sally – very realistic superb”:²²


²¹ See ElevenLabs website, “How are voices updated & changed?”, available at <https://help.elevenlabs.io/hc/en-us/articles/20939463028881-How-are-voices-updated-changed> (last verified May 31, 2024).

²² See Reddit Social Media Platform, at ElevenLabs Channel, available at https://www.reddit.com/r/ElevenLabs/comments/18jyp9z/anyone_know_how_i_can_get_bella_back/ (last verified May 31, 2024); *id.* at https://www.reddit.com/r/ElevenLabs/comments/1863dct/cant_believe_they_got_rid_of_bella_how_are_we/ (last verified May 31, 2024).




 **Think_Dance_4596** • 6mo ago


Nooooo Bella, I used her voice for edits with my fanfiction, it was the perfect voice, why didn't they remove it?! give us back the voices of Bella, Matthew and Ellie please!

  1   Reply ...




 **sammy2021sam** • 5mo ago


yes, they remove the most common used voice, i don't know why.

 1   Reply ...


 **bbthoma** • 6mo ago

Like others, I used Bella before and really liked that voice. I've gone over the other voices and don't like the sound of any others as much as I liked Bella. That seemed to be the best voice you had. Why would you remove THAT one?




 1   Reply ...


 **Regular-Rope-9286** • 6mo ago

Wow Bella was their best female voice. Not sure what to do now, maybe just go to a free voice-to-speech website.







 **AggravatingChair4835** • 6mo ago


Here is rebella: 5mWnn1lvtkl6lS5tOMEG

 1   Reply ...






 **youngdl** • 5mo ago



#BRINGBACKBELLA

  1   Reply  Award  Share ...

 **karlkablisk** • 5mo ago

We really need the word of anyone, ANYONE who has made a custom bella, has anyone tested the one, one person mentioned? rebella: 5mWnn1lvtkl6lS5tOMEG
I think we should get a few tries at this, AS A COMMUNITY.
I really can't with any other voice, its too late, I've already heard it once, I think anyone in this thread should understand... lol

 1   Reply  Award  Share ...




r/ElevenLabs • 6 mo. ago
dthings
...

Can't believe they got rid of Bella. How are we supposed to use EL with faith that a voice we're using for a major project won't be suddenly pulled?

Question


I am like 80% of the way through a project, and of course the voice of Bella gets pulled. Seems I now have to re-record the entirety of her lines with another voice (of which I can't find any that I like, so it begs the question - why'd they get rid of Bella?).

23






charlesmccarthyufc • 6mo ago

Take the lines you made and use them to clone the voice again

24

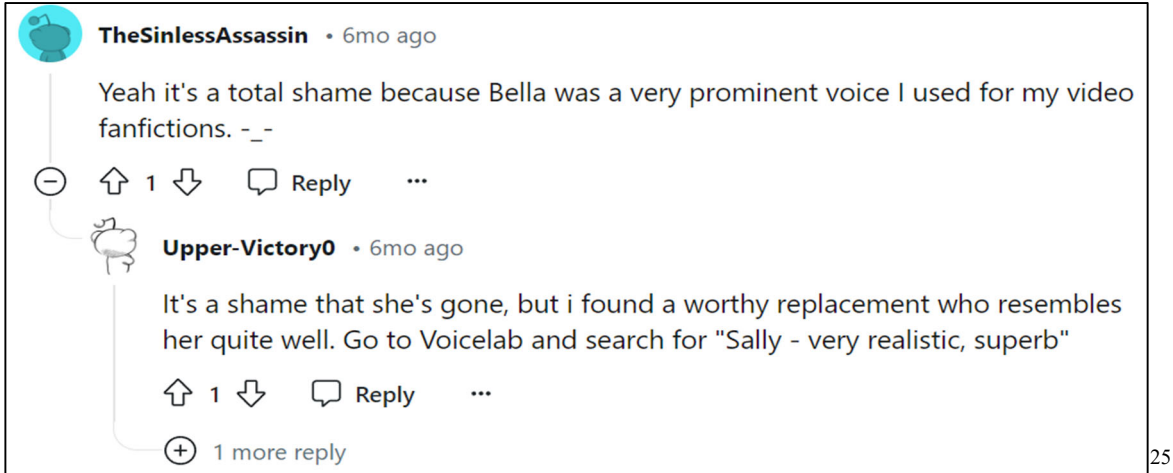

Good_Punk2 🇺🇸 • 6mo ago

I'm in the exact same situation. I recloned the voice from the recordings I have but it's not quite the same.

 7
 
 Reply
  Award
  Share
 ...

23 See https://www.reddit.com/r/ElevenLabs/comments/1863dct/cant_believe_they_got_rid_of_bella_how_are_we/ (last verified 5/31/2024). Elevenlabs Channel,

24 See https://www.reddit.com/r/ElevenLabs/comments/1863dct/cant_believe_they_got_rid_of_bella_how_are_we/ (last verified 5/31/2024). Elevenlabs Channel,



62. On information and belief, the training data ElevenLabs used to train its foundation models contains a large volume of Vacker’s audiobook narrations. Although ElevenLabs removed Bella from its website, it did not remove these recordings from its corpus of training data or retrain its foundation models after removing the “Bella” voice_id from its website and API. As a result, the unique features of Vacker’s voice remain in ElevenLabs’ system—allowing users to create high-quality realistic clones of Vacker’s voice.

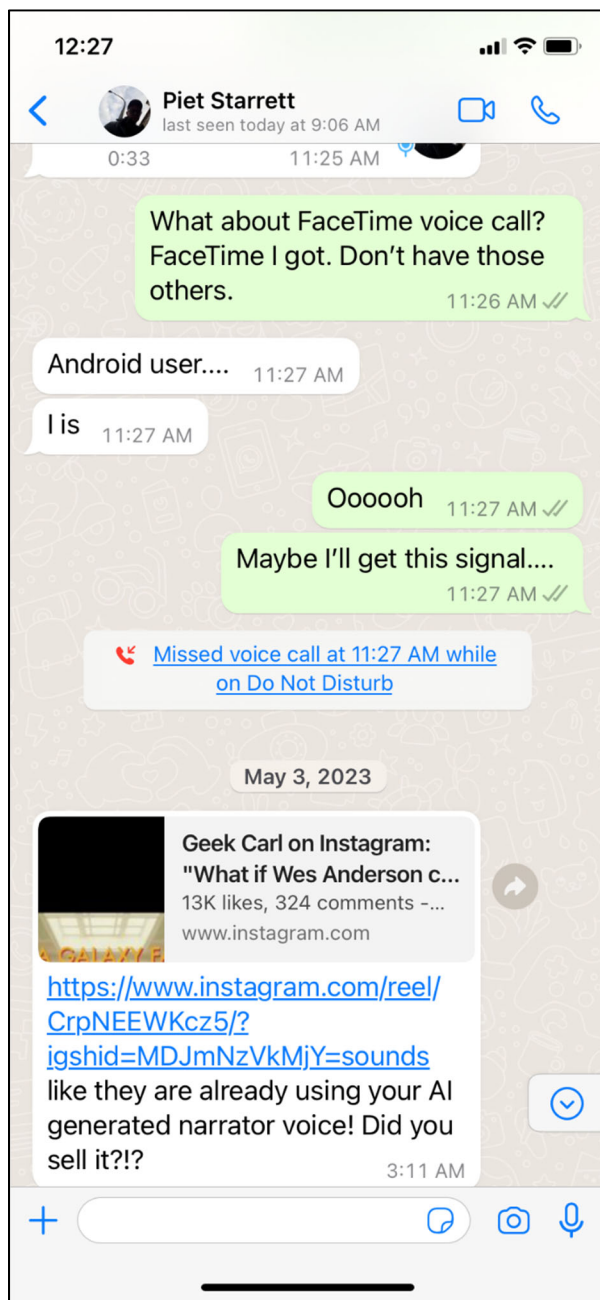
63. As a result of ElevenLabs’ actions, Vacker has suffered economic losses, damage to her professional reputation, and significant emotional distress. ElevenLabs has been unjustly enriched. Accordingly, Vacker seeks injunctive relief to prevent further misuse of her voice by ElevenLabs, damages to compensate for the harm she has already suffered, disgorgement of profits, value and/or investment achieved by ElevenLabs’ misuse, and other available monetary relief.


2. Boyett learned that ElevenLabs cloned his voice and was distributing recordings in his voice and likeness from friends and fans of his work.

64. In 2023, Boyett learned from fans that ElevenLabs was distributing recordings

²⁵ Reddit Social Media Platform, at ElevenLabs Channel, *available at* https://www.reddit.com/r/ElevenLabs/comments/18jyp9z/anyone_know_how_i_can_get_bella_back/ (last verified May 31, 2024).

performed in his voice and likeness.



 Roger Lorette ⓘ 🗑️ ★ ✉️ ! ✓

11/11/23, 11:42 AM

Neil deGrasse Tyson Warns: "Voyager 1 Has Detected 500 Unknown Objects Passing By In Space"

In this captivating YouTube video, join renowned astrophysicist Neil deGrasse Tyson as he delves into a groundbreaking revelation: Voyager 1, the intrepid sp...

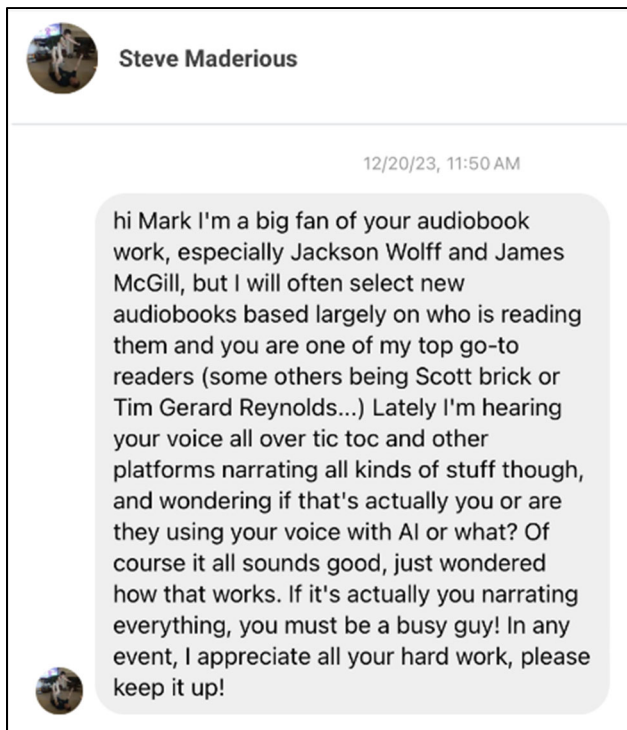
[youtube.com](https://www.youtube.com)

😊 ↩️

I've a fan of your work for a few years but lately I've heard what seems like your voice on several Youtube videos using text to speech tech. The fact that it's AI text to speech is quite obvious. I'm wondering have you licensed your voice for use by others. If this isn't your voice then it's very close. <https://www.youtube.com/watch?v=Kq6QjgQvwLY>

11/16/23, 9:49 AM

Hi Roger - Thanks so much for reaching out about this. Nope, I haven't licensed my voice for this. I've been getting similar videos from some other folks and am looking into it. Really appreciate you letting me know.



3. **ElevenLabs initially claimed their system was trained solely on two public repositories of open-source audiobooks and that its “default” voices were chosen at random.**

65. After Boyett alerted the Screen Actors Guild (“SAG”) about the problem, an attorney for SAG met with ElevenLabs in London. In September 2023, SAG updated Boyett on its conversation with ElevenLabs and forwarded him the email communications between them.

66. In emails to SAG, ElevenLabs claimed it did not use *any* copyrighted material to train its systems. ElevenLabs claimed it had trained its AI system using recordings solely from two open-source repositories of audiobook recordings, the contents of which are in the public domain—*specifically, the LibriVox and Common Voice repositories.*

67. ElevenLabs further claimed that its “Adam” default voice was “randomly generated.”

68. As discussed further below, on information and belief, ElevenLabs’ representations regarding the source material it used to train its system are false; ElevenLabs trained its AI systems

using a vast amount of high-quality audio recordings obtained from other sources—well beyond the material contained in the LibriVox and Common Voice repositories—and were well aware that the training corpus included copyrighted material. On information and belief, ElevenLabs deliberately sought out and used audiobook recordings created by professional voice actors with large libraries of audiobook recordings.

69. More specifically, without permission or consent, ElevenLabs used recordings of audiobook narrations performed by Voice-Actor Plaintiffs in the training data used to train its generative AI systems, and trained and/or selected the “Adam” and “Bella” voices based on audiobook narrations performed by Boyett and Vacker.

70. Defendant’s representations to SAG are not credible. ElevenLabs has not only created highly realistic clones of Boyett’s and Vacker’s voices that capture their unique vocal timbres, intonation, accent, prosody, pacing, stress, and delivery style, these voice clones are so similar to Voice-Actor Plaintiffs’ real voices that they have been repeatedly recognized by their fans, friends, and colleagues. Moreover, ElevenLabs’ creation of such accurate and convincing voice clones shows that ElevenLabs must have used large quantities of professional-quality audiobook recordings—including Voice-Actor Plaintiffs’ recorded audio narrations—to train its AI system.

71. ElevenLabs has not publicly identified the neural architecture of its generative AI system. However, other leading AI companies, including Google, Meta, Microsoft, and others have trained comparable text-to-speech systems with similar capabilities (including vocal cloning, text-to-speech translation, speech-to-speech translation, and multi-speaker turn-based “dubbing,”) and have publicly published research papers on their work.²⁶ Based upon similar systems,

²⁶ Le, Matthew et al., *Voicebox: Text-Guided Multilingual Universal Speech Generation at Scale*, arXiv preprint arXiv:2306.15687 (Meta, Fundamental AI Research, Oct. 19, 2023); Rubenstein, Paul K. et al., *AudioPaLM: A Large*

ElevenLabs could not—and did not—train its text-to-speech system relying solely on open-source recordings from the Common Voice and Librivox repositories, as these repositories lack sufficient quantities of professional high-quality audio narrations to train a system that can produce “voice actor-grade” text-to-speech, vocal-cloning, and speech-to-speech translation in 28 languages.

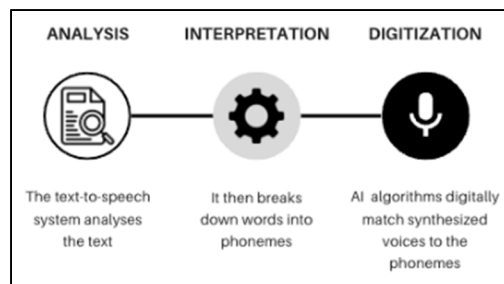
1. The ElevenLabs system uses generative AI to convert text supplied by the user to a recorded audio narration of the same text in a clone of Voice-Actor Plaintiffs’ voices.

(a) Text-To-Speech Systems—a form of Generative AI—use neural networks to convert written text to recorded speech.

72. Neural networks form the technological foundation for state-of-the-art generative AI systems across various domains, including models that generate text (like OpenAI’s GPT-4).

(b) Core components of a text-to-speech system:

73. State-of-the-art (neural) text-to-speech systems share certain core steps:²⁷



(i) Text analysis and preprocessing.

First, the system processes and interprets the input text. During this stage, the system must

Language Model That Can Speak and Listen, arXiv preprint arXiv:2306.12925 (Google, Jun. 22, 2023) (available at <https://arxiv.org/pdf/2306.15687.pdf>) (system incorporating Soundstorm below); Dong, Qianqian et al., *PolyVoice: Language Models for Speech to Speech Translation*, arXiv preprint arXiv:2306.02982v2 (ByteDance, Jun. 13, 2023); Borsos, Zalan et al., *SoundStorm: Efficient Parallel Audio Generation*, arXiv preprint arXiv:2305.09636 (Google Research, May 16, 2023); Zhang, Ziqiang et al., *Speak Foreign Languages with Your Own Voice: Cross-Lingual Neural Codec Language Modeling*, arXiv preprint arXiv:2303.03926 (Microsoft, Mar. 7, 2023) (Valle-X system); Wang, Chengyi et al., *Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers*, arXiv preprint arXiv:2301.02111 (Microsoft, Jan. 5, 2023) (monolingual Valle system).

²⁷ See ElevenLabs Website, Blog, May 1, 2023, *What is Text to Speech? (2024 Update)* (available at <https://elevenlabs.io/blog/what-is-text-to-speech/>) (narrated in Adam voice).

normalize the text (converting numbers, abbreviations, etc. to their full-word forms), identifying sentence boundaries, and understanding context and emotional cues within the text. This is crucial for generating speech that aligns with the intended message and emotional tone.

(ii) Linguistic interpretation.

The system converts the words into a sequence of graphemes (sonically significant characters, or combinations of characters), phonemes, and/or in some end-to-end systems, into semantic and acoustic “units.”²⁸ As explained by ElevenLabs, “Phonemes are the smallest units of sound in a language that can distinguish one word from another. In TTS, phonemes are crucial for accurately pronouncing words.”²⁹ The system may also identify and represent prosody (rhythm, stress, and intonation) based on the text’s context and semantic content. The resulting representations are then mapped to a multi-dimensional dense vector space called “embeddings.”³⁰ This allows the system to holistically understand the text as a sequence of phonetic sounds that are passed on to an acoustic model for speech generation.

²⁸ Some systems convert graphemes to phonemes. Some end-to-end systems can generate speech directly from graphemes without explicit phoneme conversion. Either way, the graphemes, phonemes, or semantic and acoustic “units” are mapped to dense vector representations. To capture the sequential nature of the input, positional encodings are added to these embeddings. This is essential for the model to learn the temporal dependencies in the data.

²⁹ ElevenLabs Website, Blog, May 1, 2023, *What is Text to Speech? (2024 Update)* (available at <https://elevenlabs.io/blog/what-is-text-to-speech/>) (narration available in Adam voice). Note, however, that some systems like Microsoft’s Vall-E X, and ByteDance (Tik-Tok)’s PolyVoice models directly predict acoustic speech units, rather than phonemes, that capture the phonetics and semantic content in speech.

³⁰ Depending on the architecture, the system may create an intermediate representation of the original text (phonemes, timing, stress, intonation, etc.), that it passes on, as an output to the acoustic model. The acoustic model receives this intermediate representation as an input, along with the speaker embedding, and potential information regarding timing, rhythm, stress, and intonation information derived from the text. The acoustic model then uses this information to generate Or it may do so implicitly, as the information passes through the network in an “end-to-end” system.

(iii) Acoustic model converts the phonemes into an audio performance of speech using phoneme and speaker information.

The acoustic model is able to receive as an input (i) the embeddings that represent the phoneme information,³¹ and (ii) the “speaker embedding,” which captures detailed information regarding the speaker’s voice, pitch, tone, and speaking style, and other vocal/speech traits. The acoustic model uses this information to convert the embeddings into a human-like sounding narration of the text.³² As described by ElevenLabs, “[o]nce the system phonemically interprets the text, the next step involves digitizing this speech... *AI algorithms are trained on vast datasets of spoken language, enabling them to generate speech that mimics human tonality and rhythm. This synthesized voice is then matched with the phonemes to produce speech that sounds natural.*”³³

(c) In a multi-speaker TTS with vocal cloning capabilities, the Acoustic model creates a “speaker embedding” that represents the characteristics of a voice as a point in multi-dimensional space, where each dimension correlates to some characteristic of the voice or speaking style.

74. The acoustic model is a critical component of multi-speaker text-to-speech systems with vocal cloning capabilities. To enable the acoustic model to generate speech in a variety of voices, it must be trained on a diverse dataset of speech samples from numerous speakers.

75. During this training process, the model analyzes a large dataset of speech samples from a large number of different speakers to learn the patterns and characteristics that define each unique voice. The model learns to represent these different vocal characteristics, speaking styles,

³¹ More specifically, the system receives vector embeddings that represent the phonemes, positional information, and prosodic features, such as pauses, etc., inferred from the semantic content of the text itself.

³² ElevenLabs Website, Mati Staniszewski, Blog, November 10, 2023, *What is Generative AI Audio? Everything you need to know*, (available at <https://elevenlabs.io/blog/what-is-generative-ai-audio/>).

³³ ElevenLabs Website, Blog, May 1, 2023, *What is Text to Speech? (2024 Update)* (available at <https://elevenlabs.io/blog/what-is-text-to-speech/>) (narration available in Adam voice).

and traits that vary across speakers, and represent the voice of a specific speaker as a point in a high-dimensional space called a “latent space.” In this high-dimensional latent space, each dimension represents an abstract continuum of some aspect or feature of the voices analyzed. For example, a dimension could correspond to aspects of vocal timbre and overtones, elocution, accent, pitch, rhythm, intonation, emotive styles, speaking rate, and the like.³⁴

76. Thus, each unique voice is represented as a point in a high dimensional space, where each dimension correlates to some aspect of the person’s unique voice and speaking style. Within this latent space, voices with similar characteristics are clustered closer together and dissimilar voices are placed further apart. By learning to map the relationships between these vocal traits, the acoustic model creates a comprehensive representation of the voice data.

77. The vector points within the latent space that represent the characteristics of a speaker’s voice and speaking style is called a “**speaker embedding**.” The location of this point—the speaker embedding—within the latent space identifies the unique traits of the speaker’s voice and speaking style based on all the traits learned during training.

78. During the training process, the acoustic model learns to map the input speech data to these voice embeddings in the latent space. The model adjusts the values along each dimension of the vector to minimize the difference between the generated output and the original speech data. This process allows the model to optimize representation of each voice.

79. Once the system is trained, the weights of neural network may be fixed such that the system does not learn new voices of new users during use. Nonetheless, once the latent space

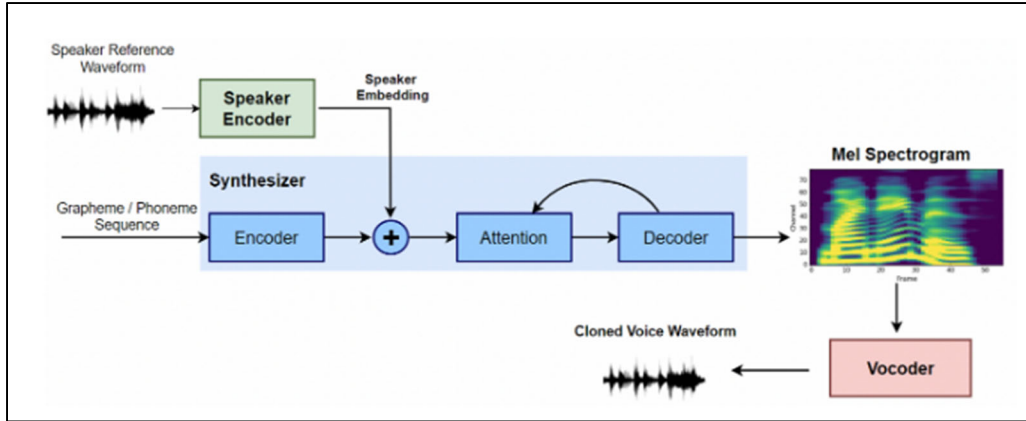
³⁴ In actuality, individual dimensions do not necessarily correspond directly to discrete, identifiable characteristics that a human would recognize, like pitch or accent. Instead, each dimension captures an abstract, learned feature that contributes to the overall representation of the voice. These learned features are often complex combinations of various aspects of the voice, making it difficult to interpret the meaning of any single dimension in isolation. See note 50, *infra*, for additional information.

is created through training, the model can “clone” the voice of a user with a high degree of realism. When a user wishes to clone a voice, the system maps the characteristics of the voice within the latent space of all known vocal characteristics it learned during training. The trained network uses an *encoder* to generate a corresponding speaker embedding vector—locating that voice within this latent space. The speaker embedding is then used to condition the text-to-speech model to generate speech that mimics the target speaker’s voice.

80. Importantly here, the more high-quality training data available for a specific speaker, the better the system will be able to reproduce that speaker’s unique vocal characteristics with a high degree of realism. Thus, the system is able to reproduce the unique voice and speaking style of a speaker for whom it has abundant training data more realistically than a speaker whose vocal traits are not well represented. High-quality data provides the model with a richer and more detailed understanding of the speaker's voice, including nuances in pitch, tone, rhythm, and speaking style, closely mimicking the target speaker and resulting in more natural and lifelike voice synthesis.

81. In the diagram below from ElevenLabs’ website: (i) the user provides the Speaker Reference recording of the voice they wish to clone; (ii) the Speaker Encoder analyzes the recording and determines where the voice would have been located in the latent space, generating the speaker embedding within the latent space; and (iii) the acoustic model receives the grapheme/phoneme information (about the text it is supposed to narrate) and the speaker embedding, and uses a decoder to generate the waveform of the audio narration within their voice, which is converted to an audio recording by a neural vocoder.³⁵

³⁵ ElevenLabs Website, Blog, Mati Staniszewski, *What is Generative AI Audio? Everything you need to know*, November 10, 2023 (*available at* <https://elevenlabs.io/blog/what-is-generative-ai-audio/>) (noting further that the generation of the audio includes both the decoder that generates the waveform, and a Discriminator that acts as a judge and rejects the generation if it does not sound sufficiently real, using a generative adversarial network architecture).



82. As ElevenLabs explains, “[b]y training these three parts with lots of speech data, our AI-based voice generator becomes a master impersonator—it understands all the nuances that make voices unique. The voices it generates are so realistic that you could easily mistake it for the real person speaking.”³⁶

83. Moreover, to allow users to modify a voice in a predictable manner (by shifting the location of the speaker embedding), ElevenLabs trained a separate model to modify the characteristics of a voice (by shifting the location of the speaker embedding within the latent space) based on more human-recognizable features: *e.g.*, gender, age, accent, pitch and speaking style. They call this interface “Voicelab.”³⁷

2. Training a state-of-the-art text-to-speech system requires a vast quantity of high-quality audio recordings of narrated text.

84. Modern text-to-speech (“TTS”) systems, which convert written text into spoken

³⁶ *Id.*

³⁷ ElevenLabs Website, Blog, ElevenLabs Team, January 11, 2023, *Enter the new year with a bang* (“We had an idea for how we’d go about this which came as we unpacked the methods we currently use for speech synthesis and voice cloning. Both processes require a way of encoding the characteristics of a particular voice. *Speaker embeddings are what carries this identity – they’re a vector representation of a speaker’s voice. We realized that we could sample from the distribution of speaker embeddings by training a dedicated model to let us create infinitely many new voices. Since our users mostly look for specific speech characteristics, we needed to add a degree of control over the process. We expanded our model with conditioning to generate voices based on their characteristics. The model now lets you set certain basic parameters which establish the new voice’s core identity: gender, age, accent, pitch and speaking style.*”).

words, rely heavily on neural networks to achieve natural-sounding speech output.³⁸ To train a TTS system to produce high-quality, natural-sounding speech output, vast quantities of high-quality audio recordings and corresponding text transcripts are required.

85. The body of content used to train a generative AI—known as the “training corpus”—typically consists of “input-output pairs,” where the input is the text and the output is the audio recording of that text being spoken. Such a model learns to generate speech by analyzing these pairs and adjusting its internal parameters to minimize the difference between its generated output and the target audio. This training approach is known as supervised learning, and it is an essential step in creating high-quality TTS systems.

86. These input-output pairs are essential for training a text-to-speech system because:

1. **Mapping text to speech:** The primary goal of a TTS system is to convert written text into natural-sounding speech. By training the system with text-speech pairs, the model learns to map the input text to the corresponding speech output. This allows the system to generate speech that accurately represents the input text.
2. **Learning pronunciation and prosody:** Text-speech pairs help the TTS system learn the correct pronunciation of words and phrases, as well as the appropriate prosody (intonation, stress, and rhythm) for different types of sentences and contexts. The model can learn these aspects by analyzing the speech output in relation to the input text.
3. **Handling diverse text:** Training on a large dataset of text-speech pairs exposes the TTS system to a wide variety of text, including different words, phrases, and sentence structures. This helps the model generalize and handle diverse text inputs, improving its overall performance and flexibility.
4. **Capturing speaker characteristics:** If the TTS system is designed to generate speech in different voices, training the system on text-speech pairs from a vast number of different speakers allows the model to learn and replicate the unique

³⁸ See ElevenLabs Website, Blog, November 10, 2023, *What is Generative AI Audio? Everything You Need to Know* (available at <https://elevenlabs.io/blog/what-is-generative-ai-audio/>) (“The development of AI is a vast field, but at a high level, deep learning and neural networks have been pivotal. These advancements enable modern AI TTS models to decipher the text, determine the appropriate intonations, and synthesize them into spoken words. This process involves training the AI with vast datasets of human speech, enabling it to generate voices that are not only indistinguishable from humans but also able to communicate feelings and nuanced meanings.”).

characteristics, such as voice quality, speaking style, and emotional tone, of each speaker.

5. **Enabling transfer learning:** Pre-training a TTS system on a large dataset of text-speech pairs allows for transfer learning. The pre-trained model can then be *fine-tuned* on a smaller dataset specific to a particular speaker, reducing the amount of data and training time required to create a specialized TTS system.

87. Thus, using input-output pairs of text and speech is crucial for training modern, state-of-the-art TTS systems, and enables the models to learn the complex mapping between text and speech, handle diverse text inputs, capture speaker characteristics, and benefit from transfer learning, ultimately resulting in high-quality, natural-sounding speech output.

88. Consistent with these benefits (and contrary to what ElevenLabs told SAG), ElevenLabs indicated in a November 2022 blog post on its website that ElevenLabs *had already trained its Speech Synthesis system on over 500,000 hours of audio*.³⁹

Both [of] our model's strengths - fluency and proper intonation - come from a wealth of training data it has seen (*over 500k hours!*), but really the central factor is how it learns from this data, which is down to the way it's built.⁴⁰

3. **Audiobook recordings are an ideal source of content for training state-of-the-art text-to-speech systems.**

89. For many reasons, audiobook narrations are prime content to train TTS systems, due to their consistent audio quality, professional narration, wide range of styles and emotions, and readily available text transcripts. Thus, the use of audiobook data for TTS training is a well-established practice in training text-to-speech systems.

90. *First*, audiobooks offer a vast amount of high-quality speech data, usually recorded in professional studio settings, ensuring clear and consistent audio quality. The controlled studio

³⁹ ElevenLabs Website, Blog, by ElevenLabs Team, November 24, 2022, *The first AI that can laugh* (available at https://elevenlabs.io/blog/the_first_ai_that_can_laugh/).

⁴⁰ *Id.*

environment maintains a uniform sound for each speaker throughout the recording, which is crucial for training TTS models to accurately capture and reproduce speaker-specific characteristics.

91. *Second*, audiobooks provide a large volume of text (typically, a full-length novel), narrated by the same speaker, in consistent, high-quality studio sound.

92. *Third*, audiobooks are typically narrated by professional voice actors who possess excellent diction, intonation, and expressiveness, providing the TTS models with examples of natural and engaging speech patterns.

93. *Fourth*, audiobooks cover a wide range of genres, styles, and emotions, exposing the models to diverse linguistic content and prosodic variations.

94. *Fifth*, audiobooks carefully follow (and are usually accompanied by) their corresponding text transcripts, which can be aligned with the audio to create text-speech pairs for training.

95. For these reasons, the use of audiobook narrations for TTS training is a well-established practice in the field, with numerous research papers and industry applications demonstrating their effectiveness in improving the naturalness, expressiveness, and overall quality of synthetic speech.

4. The training process requires making multiple copies of the audio recordings and text used to train the system.

96. The process of collecting and preparing the training corpus involves creating multiple copies of the audio recordings and text. Audio recordings must be pre-processed before they can be used to train the TTS system to normalize its format, resolution, and/or volume, while the text is tokenized and aligned with the corresponding audio segments. These pre-processing steps require the user to create new copies of the original works and are necessary to ensure that

the model can effectively learn the mapping between text and speech.

97. Audio pre-processing typically involves several steps, including: (i) converting the sampling rate of each recording to a standard sampling rate and resolution used by the system; (ii) converting stereo recordings to mono by combining the left and right audio channels; (iii) normalizing volume of the recordings; (iv) reducing and removing background noise; (v) editing out undesired silence in the recordings; and (vi) segmenting the recordings into shorter clips.

98. The pre-processing of the audio recordings requires making intermediate copies of the recordings and ultimately results in at least one pre-processed copy of each recording that is separate and distinct from the original.

99. Moreover, it is standard industry practice to maintain copies of the pre-processed and processed audio recordings, transcripts, and extracted features in a version-controlled repository and/or backup system. This results in additional copies of the pre-processed recordings and text being made each time a change is committed to the repository or a backup is created.

100. On information and belief, ElevenLabs maintains one or more copies of the pre-processed and processed audio recordings, transcripts, and extracted features used to train its system in a version-controlled repository and/or backup system.

E. The datasets ElevenLabs claims to have relied on to train its system do not contain enough high-quality validated data to train a state-of-the-art text-to-speech system with the capabilities it provides.

1. The Mozilla Common Voice project contains only 2,532 validated hours of speech in English.

101. The Mozilla Common Voice project is a crowdsourcing project started in 2019 by Mozilla to create a free opensource repository of recordings to train *speech recognition* software.⁴¹

⁴¹ See Mozilla Common Voice Website, “About” page available at <https://commonvoice.mozilla.org/en/about> at Wikipedia article on Common Voice, available at

Individuals contribute to the project by donating short audio recordings of themselves narrating text. Listeners also may contribute by validating the audio recordings to ensure quality.⁴² Mozilla updates the database with new validated entries every three months, making the software available for download to anyone who wishes to use it.⁴³ In September 2023, the Common Voice English dataset (Common Voice Corpus 15.0, released 9/13/2023) contained a **total of 3,347** recorded hours of speech, of which a total of **2,532** hours had been *validated*.

102. The Common Voice repository contains only very short recordings from a given speaker recorded under consistent conditions, and is verified only to meet the quality necessary to train a speech-to-text system. In contrast, in a text-to-speech system, the system uses the *audio recording* as a model for what the system should output based on the text as input—making the recording and narration quality of utmost importance.⁴⁴

2. The Librivox repository contains unvalidated recordings of varying quality.

103. The Librivox repository is a voluminous collection of audiobook recordings created by volunteers who read public domain texts aloud.⁴⁵ The repository is near-exclusively comprised of narrations of government documents and private works written prior to 1929.⁴⁶

104. This repository has significant limitations when it comes to training modern, high-

https://en.wikipedia.org/wiki/Common_Voice#:~:text=Common%20Voice%20is%20a%20crowdsourcing,review%20recordings%20of%20other%20users.) (last verified, April 2, 2024).

⁴² See Mozilla Common Voice Website, “About” page, (available at <https://commonvoice.mozilla.org/en/about>) Wikipedia article on Common Voice, available at https://en.wikipedia.org/wiki/Common_Voice#:~:text=Common%20Voice%20is%20a%20crowdsourcing,review%20recordings%20of%20other%20users.) (last verified, April 2, 2024).

⁴³ See Mozilla Common Voice Website, “Download the Dataset” page, (available at <https://commonvoice.mozilla.org/en/datasets>) (last verified, August 1, 2024)

⁴⁴ Heiga Zen, et. al., *LibriTTS: A Corpus Derived from LibriSpeech for Text-to-Speech* (2019) (explaining why corpus designed for speech-to-text was inadequate to train a text-to-speech system).

⁴⁵ See LibriVox Website (available at <https://librivox.org/>) (last verified April 2, 2024).

⁴⁶ *Id.*

quality text-to-speech systems, including low sample rate recordings, low bit-depth recordings, use of poor-quality microphones, poor signal-to-noise ratios, suboptimal ambient recording conditions, highly inconsistent narration quality,⁴⁷ unclear enunciation, and inappropriate style and delivery.⁴⁸

105. As a result, training a state-of-the-art text-to-speech system solely on the Librivox repository, even with pre-processing, will not yield the high-quality, natural-sounding output that ElevenLabs' AI system produces.

106. Even setting aside these issues, the Librivox repository does not contain sufficient high-quality audio to train a system with the capabilities of the ElevenLabs system.

107. By August 2023, ElevenLabs' Multilingual v2 model could convert text to speech using its "default" voices, including "Adam" and "Bella," **in 28 languages**—including French, German, Hindi, Italian, Polish, Portuguese, and Spanish,⁴⁹ as well as Chinese, ***Korean***, Dutch, ***Turkish***, Swedish, Indonesian, ***Filipino***, Japanese, Ukrainian, Greek, ***Czech***, Finnish, Romanian, Danish, Bulgarian, ***Maylay***, ***Slovak***, ***Croatian***, Classic Arabic, and Tamil.⁵⁰

108. As ElevenLabs readily admits on its website, "[t]he quality of TTS in any language

⁴⁷ See, *LibriS2S: A German-English Speech-to-Speech Translation Corpus*, Pedro Jeuris and Jan Niehues, Maastricht University (April 22, 2022) (discussing advantages and disadvantages of Librivox dataset) (available at <https://arxiv.org/pdf/2204.10593.pdf>) (e.g., "The speakers that record the audio books on librivox do not always have access to a high-quality microphone. Because of this it can be challenging to create a TTS system for a single speaker."); Learning from Flawed Data: weakly supervised automatic speech recognition, Dongji Gao, Johns Hopkins University, Nvidia, and Xiaomi Corp. (September 26, 2023), (available at <https://arxiv.org/pdf/2309.15796.pdf>) (last verified April 2, 2024).

⁴⁸ Dongji Gao, Johns Hopkins University, Nvidia, and Xiaomi Corp. *Learning from Flawed Data: weakly supervised automatic speech recognition* (September 26, 2023), (available at <https://arxiv.org/pdf/2309.15796.pdf>) (last verified April 2, 2024) (addressing problems with Librivox data *even for speech-to-text* training).

⁴⁹ ElevenLabs Website, Blog, April 27, 2023, by ElevenLabs Team, *Introducing ElevenLabs Multilingual v1: Our New Speech Synthesis Model*, (available at <https://elevenlabs.io/blog/eleven-multilingual-v1/>) (with article narration in "Adam" voice).

⁵⁰ See ElevenLabs Website, Blog Entry, August 22, 2023, by ElevenLabs Team Multilingual v2 (*available at* <https://elevenlabs.io/blog/multilingualv2/>).

depends on the depth of the dataset it's trained on and the sophistication of the algorithms used.”⁵¹

Yet, the Librivox repository has a total of:

- Only **5** short works in Korean—two of which are only partly in Korean;
- Only **6** books in Czech;
- Only **2** books in Slovak;⁵²
- Only **5** books in Turkish;
- Only **5** books in Croatian;
- Only **7** books in Tamil;
- Only **11** books in Bulgarian; and
- **Zero** recordings in Filipino.⁵³

109. In short, on information and belief, ElevenLabs' claim in September 2023 that it had trained its text-to-speech model solely using data from the Common Voice and Librivox repositories is false.

110. In comparison, Microsoft's VALLE-X system released in 2023, which can convert text to speech only in English and Chinese, was trained on 60,000 hours of high-quality audiobook recordings in English, and 10,000 hours of recordings in Chinese, as well as the MT data from AI Challenger5, OpenSubtitles20186 and WMT20207, which contain about 13M, 10M, and 50M text-speech pairs, respectively.⁵⁴

⁵¹ May 1, 2023, ElevenLabs Website, Blog, Mati Staniszewski (*available at* <https://elevenlabs.io/blog/what-is-text-to-speech/>).

⁵² Likewise, the Common Voice Corpus has only 27 hours of audio recorded, with a total of 22 validated hours—likely comprised of the same material.

⁵³ See Mozilla Common Voice website, at Datasets (under Language tab) (*available at* <https://commonvoice.mozilla.org/en/datasets>).

⁵⁴ Microsoft, Speak Foreign Languages with Your Own Voice: Cross-Lingual Neural Codec Language Modeling (*available at* <https://arxiv.org/pdf/2303.03926.pdf>).

111. The ByteDance (Tik-Tok) Polyvoice system was trained on 140,000 hours of speech.⁵⁵

112. Meta’s Voicebox multilingual text-to-speech system has vocal cloning capabilities and can synthesize speech in English, French, German, Spanish, Polish and Portuguese.⁵⁶ It was trained on 60,000 high quality hours of English audio and 50,000 hours of multi-lingual audio (French, German, Spanish, Polish, and Portuguese).⁵⁷

113. Google’s Soundstorm acoustic model was trained on 60,000 hours of English speech data from the libri-light corpus to generate acoustic features from semantic tokens.⁵⁸ However, to enable the full text-to-speech system, it requires a separate text-to-semantic transformer model trained to map text to the semantic token inputs which was trained on a separate dataset of 100,000 additional (distinct) hours of English-language dialogues.⁵⁹

114. The quantity of high-quality audio data required to train a state-of-the-art text-to-speech system with the capabilities of the ElevenLabs system—whose system includes *text-to-speech, voice cloning, and multi-lingual voice-to-voice translation capabilities in 29 languages*—vastly exceeds the quantity of such content available within the LibriVox and Common Voice repositories, and required ElevenLabs to scrape, download, or otherwise obtain copyrighted audio recordings and use them without permission to train its AI systems.

115. Indeed, by its own admission—*prior* to releasing its first monolingual (English

⁵⁵ ByteDance, Polyvoice, <https://arxiv.org/pdf/2306.02982.pdf> (2023) (speech to speech translation model).

⁵⁶ See Meta, Matthew Le et al., *Voicebox: Text-Guided Multiilingual Universal Speech Generation at Scale*, (June 23, 2023) (available at <https://arxiv.org/pdf/2306.15687.pdf>).

⁵⁷ See Meta, Matthew Le et al., *Voicebox: Text-Guided Multiilingual Universal Speech Generation at Scale*, (June 23, 2023) (available at <https://arxiv.org/pdf/2306.15687.pdf>).

⁵⁸ Zalan Borsos, et al. Google, *Soundstorm: Efficient Parallel Audio Generation* (May 16, 2023) (available at <https://arxiv.org/pdf/2305.09636.pdf>).

⁵⁹ Zalan Borsos, et al. Google, *Soundstorm: Efficient Parallel Audio Generation* (May 16, 2023) (available at <https://arxiv.org/pdf/2305.09636.pdf>).

only) Speech Synthesis model—ElevenLabs had already trained its Speech Synthesis model on **500,000 hours** of audio.⁶⁰

116. On information and belief, ElevenLabs obtained the recordings for its training corpus from sources including the open internet, “torrent” file-sharing sites, and/or by creating unauthorized copies of audio recordings distributed legally through audiobook subscription services like Amazon’s Audible service.

F. ElevenLabs’ misappropriation of the Voice-Actor Plaintiffs’ voices and likenesses causes irreparable injury that cannot be quantified.

1. Injury to the Voice-Actor Plaintiffs

117. A professional narrator brings his or her unique vocal presence, character, and speaking style to bear on written works to craft clear and emotionally resonant audio narrations that engage and delight audiences.

118. A professional voice actor develops his or her unique vocal presence, character, and speaking style, through years of training, practice, and refinement.

119. A professional voice actor’s unique vocal presence and speaking style is inseparable from his individual identity; those who hear it recognize him or her as the speaker.

120. A professional voice actor’s unique voice and style is likewise inseparably associated with his body of work and professional identity as an artist. The recordings of a professional voice actor’s performances define his or her character, style, and identity as an artist to the public.

121. For this reason, voice actors must curate the projects with which he or she is associated. The roles they accept and the content they narrate become intrinsically linked to their

⁶⁰ ElevenLabs Website, Blog, by ElevenLabs Team, November 24, 2022, *The first AI that can laugh* (available at https://elevenlabs.io/blog/the_first_ai_that_can_laugh/).

professional identity. The audience develops an association between the voice actor's voice and the various characters or narratives he or she has brought to life. By consistently engaging in high-quality, on-brand work, voice actors cultivate a reputation for excellence and reliability, ensuring that their voice remains a valuable presence in the industry.

122. The "Adam" voice is substantially similar to and recognizable by others as Boyett's voice.

123. The "Bella" voice is substantially similar to and recognizable by others as Vacker's voice.

124. ElevenLabs has broadcast and performed on its website some unknown number of audio recordings narrated in its default "Adam" and "Bella" default voices, inviting any visitor to its website to enter text and obtain a professional voice actor narration using its "Adam" and "Bella" voice clones.

125. Moreover, ElevenLabs grants its users unrestricted full rights to download the synthetic audio narrations they created and use them for any commercial purpose, enabling users to copy, distribute, sell, and/or otherwise use and disseminate the audio narration for commercial purposes in their own projects.

126. As a result, Vacker's and Boyett's voices and identities are being utilized in a vast assortment of media, including low-budget or irreputable projects, to promote messaging over which they have no control.

i. Injury to Boyett

127. ElevenLabs' use of Boyett's voice and style has caused and continues to cause him irreparable harm and injury, including:

a. Misappropriating his likeness, identity, and infringing his publicity rights by creating audio narrations in his voice, style, and likeness, without his consent or

compensation, and performing them on its website to promote its text-to-speech service to attract new users;

b. Distributing recordings of audio narrations created in Boyett's voice, style, and likeness to users to use as they wish, thereby implying Boyett's endorsement or affiliation with not only ElevenLabs' service but the projects and messaging of its users. As a result, Boyett's unique voice and style is used without his consent in projects with which he has no affiliation to promote messages over which he has no control, thereby damaging and diluting Boyett's reputation and credibility as a voice actor;

c. Engaging in unfair competition and false advertising by deceiving or misleading consumers or potential consumers by using Boyett's voice—which is associated with his reputation and goodwill—to promote or sell its service, and by creating a likelihood of confusion or deception to the public;

d. Diminishing Boyett's marketability and value as a voice actor by saturating the market with his voice and creating a virtually free substitute for his services; and

e. Causing Boyett emotional distress and anguish by violating his privacy and dignity and by exploiting his voice for ElevenLabs' benefit.

ii. Injury to Vacker

128. ElevenLabs' use of Vacker's voice has caused and continues to cause her irreparable harm and injury, including:

a. Misappropriating Vacker's likeness and identity and infringing her publicity rights by using her voice without her consent or compensation to promote its text-to-speech service to attract new users;

b. Implying Vacker's endorsement or affiliation with not only ElevenLabs' service—but the projects and content created by its users. As a result, Vacker's unique

voice and likeness is used without her consent, associated with projects with which she has no affiliation, promotes messages over which she has no control, damages and dilutes Vacker's reputation and credibility as a voice actor;

c. Engaging in unfair competition and false advertising by deceiving or misleading consumers using Vacker's voice which is associated with her reputation and goodwill to promote or sell its service, and by creating a likelihood of confusion or deception;

d. Diminishing Vacker's marketability and value as a voice actor by saturating the market with her voice and by creating a virtually free substitute for her services; and

e. Causing Vacker emotional distress and anguish by violating her privacy and dignity and by exploiting her voice for ElevenLabs' benefit.

129. On information and belief, ElevenLabs *knowingly* used recordings narrated by Vacker and Boyett—both successful professional voice actors who gain their livelihood by delighting audiences with their narrations of text—to train or to select the default ElevenLabs voices “Bella” and “Adam,” respectively.

CLAIMS FOR RELIEF

A. Infringement of Personality Rights, Including the Right to Privacy, Rights of Publicity, and Misappropriation of Likeness or Identity.

1. Invasion of Privacy, through the Misappropriation of Vacker's Likeness and Right of Publicity, under Texas Common Law

130. Plaintiffs incorporate by reference the allegations in paragraphs 1 through 129 as if fully set forth herein.

131. Vacker's voice, character and style of vocal delivery is a distinctive and recognizable aspect of her identity.

132. Vacker's voice is also highly valuable. Indeed, Vacker earns her livelihood by

charging fees for her services narrating text in her distinctive voice and vocal style.

133. ElevenLabs used Vacker's voice, or a copy thereof, without Vacker's authorization, consent, or compensation. Vacker learned that ElevenLabs had 'cloned' her voice and was delivering recordings in her unique and recognizable voice and likeness through third parties.

134. ElevenLabs used Vacker's voice and likeness for commercial purposes, including:

a. To promote its own services by offering at least 1 million new potential customers the free use of the "Bella" voice, resulting in ElevenLabs gaining over 1 million registered users in the first six months of being launched, and some unknown number of registered users thereafter; and

b. Distributing audio files of text narrated in the voice and likeness of Vacker to its users in exchange for subscription fees and usage-based fees for the use of its text-to-speech services accessed through its website and API.

135. Even after ElevenLabs removed "Bella" from its website, on information and belief, ElevenLabs continued to allow its API users to create and market synthetic audio narrations using the "Bella" voice rendered in the voice and likeness of Vacker, and has allowed users to recreate substantially similar voices to "Bella" and Vacker.

136. ElevenLabs' use of Vacker's voice has caused and continues to cause harm and confusion to Vacker, by implying her endorsement or affiliation with ElevenLabs' service, violating her privacy and dignity, and exploiting her voice for ElevenLabs' benefit.

137. ElevenLabs' use of the "Bella" (or substantially similar) voice and its decision to allow users, in exchange for compensation, to download audio recordings of text narrated in the unique likeness of Vacker's voice, accent, timbre and style has caused and continues to cause harm and confusion to Vacker, by implying Vacker's endorsement or affiliation with the projects,

services, and messaging of ElevenLabs' users; by saturating the market with Vacker's voice, diminishing her brand and demand for her services; and creating an unauthorized and unlicensed alternative for her services.

138. ElevenLabs' acts constitute a willful and intentional misappropriation of Vacker's likeness, identity, and an infringement of her publicity rights, in violation of the common law.

139. ElevenLabs has profited from its misappropriation of Vacker's voice and talents—enticing users to join its service, based in part on the quality of Vacker's narrations and identifiable voice and style—by gaining millions of new users delighted by the narrations they “generated,” and by securing over \$100 million in new investment, helping to increase the estimated company valuation to over \$1 billion.

140. Moreover, Vacker has suffered and continues to suffer actual damages, including lost profits and royalties, in an amount to be proven at trial, and is entitled to statutory damages, in an amount to be determined by the Court, pursuant to applicable laws.

141. Vacker also is entitled to recover other monetary relief, including the advantage or benefit derived from ElevenLabs' unauthorized use of her likeness.

142. Vacker is also entitled to injunctive relief, preventing ElevenLabs from further misappropriating Vacker's likeness or identity through the distribution of audio narrations delivered in her unique voice, style and likeness, requiring ElevenLabs to remove or delete any content or materials that use or feature the “Bella” (or substantially similar) voice on ElevenLabs' website or any other media or platform, instructing its users to do the same, and take steps to prevent users from cloning Vacker's voice using ElevenLabs' system.

b. Unjust Enrichment under Texas Law

143. Vacker's unique voice and vocal style are highly valuable assets that are inextricably linked to her personal and professional identity as a voice actor. ElevenLabs, without

Vacker's authorization or consent, has misappropriated her voice and identity by creating a voice clone named "Bella" (and allowing its users to create substantially similar voice clones) that closely matches Vacker's distinctive vocal timbre, accent, intonation, pacing, vocal mannerisms, and speaking style. This voice clone was used to promote ElevenLabs' services, attract millions of users, and generate significant revenue, market value, and investment for the company. By using this voice clone without Vacker's permission, ElevenLabs has profited from her valuable assets and caused harm to her professional reputation and potential income.

144. As a direct result of ElevenLabs' actions, Vacker has suffered economic losses, damage to her professional reputation, and significant emotional distress. Vacker seeks injunctive relief to prevent further misuse of her voice by ElevenLabs, damages to compensate for the harm she has already suffered, and other available monetary relief.

b. Misappropriation of Boyett's Likeness and Publicity Rights Under New York Civil Rights Law § 51

145. Plaintiffs incorporate by reference the allegations in paragraphs 1 through 144 as if fully set forth herein.

146. Boyett's voice is a distinctive and recognizable aspect of his identity, and the very source of his livelihood. This right is protected by the common law and statutory law of New York and other states.

147. ElevenLabs has used Boyett's voice, or a copy or imitation thereof, without Boyett's authorization, consent, or compensation. Indeed, Boyett learned that ElevenLabs had 'cloned' his voice and was delivering recordings in his unique voice and likeness through third parties.

148. ElevenLabs is using Boyett's voice and likeness for commercial purposes, including:

- a. To promote its own services by offering at least 1 million new potential customers the free use of the “Adam” voice, resulting in ElevenLabs gaining over 1 million registered users in the first six months after launching, and some unknown number of registered users thereafter; and
- b. To generate subscription and usage-based fees for the use of its text-to-speech services accessed through its website and API.

149. ElevenLabs continues to promote the “Adam” voice on its website and allow users to create synthetic audio narrations in the voice and likeness of “Adam.”

150. Defendant’s use of the “Adam” voice—and more specifically, its decision to allow users, in exchange for compensation, to download audio recordings of text narrated in the unique likeness of Boyett’s voice, accent, timbre and style—has caused and continues to cause harm and confusion to Boyett by implying his endorsement of or affiliation with the projects, services, and messaging of ElevenLabs’ users; by saturating the market with his voice, diminishing his brand and demand for his services; and creating an unauthorized and unlicensed alternative for his services.

151. Narrations in the “Adam” voice and its progeny are being created and delivered to users in the state of New York.

152. ElevenLabs’ acts constitute willful and intentional misappropriation of Boyett’s likeness or identity, in violation of the common law and statutory law of New York.

**B. Claims for Violation of the Digital Millennium Copyright Act
(Author/Publisher Plaintiffs)**

1. Violation of DMCA Anticircumvention Provisions, 17 U.S.C. §§ 1201 and 1203

153. Author/Publisher Plaintiffs incorporate by reference the allegations in paragraphs 1 through 152 as if fully set forth herein.

154. The DMCA provides two important protections for copyrighted works: (1) Section 1201 prohibits the circumvention of technological measures that effectively control access to a copyrighted work, and (2) Section 1202 prohibits the removal or alteration of copyright management information, including watermarks and copyright ownership information.

155. As discussed above, ElevenLabs has used a substantial number of Author/Publisher Plaintiffs' copyrighted audio recordings to train its AI system, as achieving the system's level of realism and authenticity in the "Bella" and "Adam" default voices required a significant quantity of high-quality training data of Boyett's and Vacker's specific recordings. Contrary to ElevenLabs' claims, ElevenLabs could not have randomly generated hundreds of voices from a near-infinite set of possibilities, and by dumb luck generated two voices recognizable in all aspects to two professional voice actors specializing in audiobook narration.

2. Violation of DMCA § 1201(a)(1)(A) (The Author/Publisher Plaintiffs)

156. Under DMCA § 1201(a)(1)(A), "No person shall circumvent a technological measure that effectively controls access to a work protected under this title." This provision is designed to protect copyrighted works from unauthorized access by prohibiting the circumvention of technological measures, such as encryption and digital rights management ("DRM") technologies.

157. Author/Publisher Plaintiffs' audiobook narrations are legally distributed via the internet exclusively through digital files protected by technological measures, including encryption and DRM technologies, which effectively control access to these copyrighted works and prevent unauthorized copying through technological means. Additionally, these files contain copyright management information, such as watermarks and metadata, that identify the work, the author, and the copyright owner.

158. For ElevenLabs to use Author/Publisher Plaintiffs' professional audiobook

narrations, which are distributed for profit by Amazon and other online retailers in DRM-protected formats, to train its text-to-speech system ElevenLabs must first circumvent the DRM protections. These DRM protections control access to the works by ensuring that only licensed copies of the works are readable, using technological means to prevent unauthorized copying or modification of these files or their contents—including the extraction and conversion of their contents into unprotected formats. This process would involve decrypting the files and removing the DRM protections, in violation of Section 1201(a)(1) of the DMCA.

159. The process of preparing the audio files for use in training a neural network text-to-speech system like ElevenLabs' speech synthesis system requires several transformations that necessitate the creation of additional unauthorized copies. These transformations require at least some—and potentially all—of the following steps: (a) converting each audio recording to a recording with a standard sampling rate (and likely, standard bit-depth and encoding format); (b) segmenting the audio recording into segments of shorter duration; (c) processing the audio recording to remove background noise; (d) normalizing the volume of individual segments; and (e) removing long pauses (silence) at the beginning, end, and/or within the recording.

160. The DRM protections on Author/Publisher Plaintiffs' audiobook narrations prevent the creation of such copies or any modification of the original files without authorization.

161. Circumventing these DRM protections is necessary to enable ElevenLabs to ingest the protected works into its AI training system.

162. By circumventing these DRM protections to create the necessary copies for its AI training pipeline, ElevenLabs violated Section 1201(a)(1) of the DMCA.

Violation of DMCA § 1201(a)(2)

163. Section 1201(a)(2) prohibits the manufacture of any tools or services designed to circumvent technological protection measures. Author/Publisher Plaintiffs' literary works and

audiobook narrations are distributed through Amazon and other online retailers in file formats subject to technological protection measures. Specifically, the DRM that protects Author/Publisher Plaintiffs' works uses technological measures to prevent unauthorized persons from opening the file, copying, modifying, or converting its contents into an unprotected format.

164. On information and belief, ElevenLabs created data processing procedures and/or processes to streamline the removal of DRM protections from DRM-protected works (including but not limited to e-books and professionally narrated audiobook recordings distributed through online marketplaces and/or subscription services) and convert the audio and/or text to an unprotected format that can be copied, segmented, and used to train its AI models, in violation of § 1201(a)(2).

Violation of 17 U.S.C. § 1202(b) (The Author/Publisher Plaintiffs)

165. Section 1202(b) prohibits the intentional removal or alteration of copyright management information, which includes identifying information about the work, the author, the copyright owner, and the terms and conditions for use of the work. This provision helps to ensure the integrity of copyright ownership information and prevent the unauthorized use of copyrighted works.

166. The audiobook narrations performed by Boyett for the Author/Publisher Plaintiffs contain Copyright Management Information ("CMI"), including metadata identifying the work, the author, the narrator, the copyright owner, and other related information. This CMI is conveyed in connection with the narrations and is protected under 17 U.S.C. § 1202(c).

167. On information and belief, to use the Author/Publisher Plaintiffs' audiobook narrations as training data for its AI system, ElevenLabs intentionally removed or altered the CMI from the digital audio files.

168. The removal or alteration of the CMI was necessary to enable ElevenLabs to ingest

the audio files into its AI training system.

169. Watermarks do not contribute to the linguistic or acoustic information necessary for training a TTS system. The CMI does not contain any phonetic information relevant to the AI training process and would likely confuse or corrupt the system if not removed. In fact, the presence of watermarks can interfere with the training process in several ways:

- a. **Noise introduction:** Audible watermarks can be perceived as noise or distortions in the audio signal. This extra noise can confuse the AI model during training, as it may try to learn patterns from the watermark rather than the actual speech content.
- b. **Inconsistency:** Watermarks may not be present uniformly throughout the audio recordings. This inconsistency can lead to the AI model learning incorrect associations between the text and the acoustic features.
- c. **Lack of relevance:** Watermarks do not carry any meaningful information related to the text or the speech itself. Including them in the training data would not contribute to the AI model's ability to generate human-like speech from input text.
- d. **Potential bias and/or interference:** If watermarks are present in some recordings but not others, the AI model might learn to associate certain acoustic features with the presence or absence of watermarks, introducing bias into the generated speech and/or otherwise interfere with training. To ensure the effective training of a TTS system, it is essential to use clean, high-quality speech data that accurately represents the relationship between the text and the corresponding acoustic features.

170. Removing watermarks from the training data helps to eliminate potential sources of noise, inconsistency, and bias, allowing the AI model to focus on learning the relevant patterns

for generating human-like speech.

171. ElevenLabs engaged in these acts knowing, or having reasonable grounds to know, that doing so would induce, enable, facilitate or conceal an infringement of copyright in the audiobook narrations.

172. ElevenLabs' conduct violates 17 U.S.C. § 1202(b)(1) and (3).

173. The Author/Publisher Plaintiffs have suffered actual damages as a result of ElevenLabs' violations of Section 1202 of the DMCA, in the form of lost licensing fees.

174. Pursuant to 17 U.S.C. § 1203(c), the Author/Publisher Plaintiffs are entitled to recover from ElevenLabs the actual damages they suffered due to ElevenLabs' violations of § 1202 of the DMCA, and any profits of ElevenLabs attributable to the violations not taken into account in computing actual damages, or, at the Author/Publisher Plaintiffs' election, statutory damages.

175. ElevenLabs' conduct has caused irreparable harm to the Author/Publisher Plaintiffs, and unless restrained and enjoined, will continue to cause irreparable harm to the Author/Publisher Plaintiffs, by creating unlicensed and unauthorized copies of their work that are free of DRM and copyright information, and risking their inclusion in future datasets used by ElevenLabs, its employees, and/or distributed to others for use in AI training.

176. The Author/Publisher Plaintiffs are entitled to injunctive relief to prevent ElevenLabs from engaging in further violations of Sections 1201 and 1202 of the DMCA pursuant to 17 U.S.C. § 1203(b).

177. On information and belief, ElevenLabs circumvented the technological measures protecting the Author/Publisher Plaintiffs' audiobook narrations and removed or altered the CMI associated with these files, without authorization, to train its AI systems.

178. ElevenLabs' actions constitute violations of both Section 1201(a)(1)(A) and Section 1202(b) of the DMCA.

179. Under Section 1203 of the DMCA, the Author/Publisher Plaintiffs, as the injured parties, are entitled to injunctive relief, actual damages, the recovery of any ill-gotten gains, statutory damages, and attorney's fees and costs.

PRAYER FOR RELIEF

WHEREFORE, Plaintiffs pray for the following relief:

(A) A permanent injunction enjoining ElevenLabs from creating and distributing to its users any synthetic audio recordings of audio narrations generated in the likeness of Vacker's or Boyett's voices, without permission, and requiring ElevenLabs to (i) instruct its users not to create further recordings using "Bella," "Adam," or any other substantially similar voice clone, and (ii) take other reasonable and necessary steps to prevent further use of Bella, Adam, or substantially similar voice clones;

(B) An award to the Voice-Actor Plaintiffs, for damages and other monetary relief, as a result of ElevenLabs' infringement of their publicity rights and misappropriation of their voices and likenesses for commercial gain, including but not limited to ElevenLabs' ill-gotten gains, Voice-Actor Plaintiff's lost profits, or both—and no less than a reasonable royalty together with interest and costs—in an amount to be determined at trial;

(C) An award to the Author/Publisher Plaintiffs for damages and other monetary relief, as a result of ElevenLabs' violation of DMCA's Digital Anticircumvention Provisions, under Sections 1201-1203;

(D) An award to the Voice-Actor Plaintiffs for exemplary damages under N.Y. Civ. Rights Law § 51 and Texas law, in an amount to be determined at trial, for the knowing use of such Voice-Actor Plaintiffs' voices and likenesses without authorization;

- (E) That this Court assess pre-judgment and post-judgment interest and costs;
- (F) That this Court award reasonable attorneys' fees; and
- (G) Such other and further relief as this Court may deem just and proper.

DEMAND FOR A JURY TRIAL

Plaintiffs demand a trial by jury of all matters to which it is entitled pursuant to Federal Rule of Civil Procedure 38.

Dated: August 29, 2024

Respectfully submitted,

OF COUNSEL:

Michael C. Wilson
Charles Theodore Zerner
Abigail R. Karol
MUNCK WILSON MANDALA, LLP
600 Banner Place Tower
12770 Coit Road
Dallas, TX 75251
(972) 628-3600

FARNAN LLP

/s/ Michael J. Farnan
Brian E. Farnan (Bar No. 4089)
Michael J. Farnan (Bar No. 5165)
919 North Market Street, 12th Floor
Wilmington, DE 19801
(302) 777-0300 (Telephone)
(302) 777-0301 (Facsimile)
bfarnan@farnanlaw.com
mfarnan@farnanlaw.com

Attorneys for Plaintiffs Karissa Vacker, Mark Boyett, Brian Larson, Iron Tower Press, Inc., and Vaughn Heppner