



## Synthetic Voice MESSAGING SUMMARY

Synthetic Voice Study Group  
Open Voice Network of the Linux Foundation  
2022.01.24  
FINAL / Isham

Synthetic voice is a powerful tool and as such, must be protected against misuse and exploitation of the technology.

### 30 Second Elevator Speech

Synthetic voice is a machine generated replica of a natural human voice. When used properly, synthetic voice can increase access and inclusion, add layers of protectivity to sensitive documents and information, and improve customer service. When used improperly, synthetic voice can cause data breaches, and identity and intellectual property theft.

### Executive Summary

As many would associate synthetic voice in a negative light, the OVON is working hard to educate the public and other stakeholders that it is more than just deep fakes and misleading information. Synthetic voice has incredible potential to serve in a positive and beneficial way. Knowing the difference between those good and bad implications is crucial as society moves forward with this fast-moving technology.

### Industry Size / Background Data

According to the Edison Research Report Infinite Dial 2021<sup>1</sup>, in February, 193 million people in the US consumed online audio. In fact, according to Beyond Words (aka SpeechKit), when new users engage with AI audio content on a website, their average time spent is 5 minutes and 53 seconds as opposed to the 30 second norm. Those users also visit 19% more pages per session.<sup>2</sup>

The number of engaged listeners in spoken word audio is on the rise. As of 2021, 22 million more people are listening to it than 7 years prior.<sup>3</sup> 127 million people daily consume audio content as opposed to 105



in 2014. And according to Edison Research Share of Ear, time spent is also increasing: a 71% increase in the last seven years by females and 23% in males.<sup>4</sup>

The important takeaway is that more people are adopting and utilizing voice in their media and information consumption. And that use is accounting for more and more of their time spent.

### What Is a Synthetic Voice?

A synthetic voice is a machine generated replica of a natural human voice. The most popular form of creating a synthetic voice (speech synthesis) is when the human voice “teaches” the machine through spoken data and corresponding transcripts how to learn and predict its words and intonations. Once trained, the technology can transform novel text into speech with the pronunciation and intonation of the natural human voice associated. This allows for the human voice to not have to repeat words repeatedly as the machine has learned and banked multiple iterations.<sup>5</sup>

### What Is the Market Value for Synthetic Voice?

As of 2020, the valuation of the Global Voice Recognition Market projected to be \$10.7 billion (USD). This sum is expected to grow to \$27.16 billion by 2026.<sup>6</sup> The primary use case of voice tech is voice assistance and in 2020, there were 4.2 billion voice assistants being used around the world. That number is expected to balloon to 8.4 billion by 2024.<sup>7</sup>

According to Report Linker, the global voice cloning market is expected to reach \$3.8 billion by 2027<sup>8</sup>. Meticulous Researcher says the speech and voice recognition market will reach \$26.8 billion by 2025<sup>9</sup>.

Text to speech technology is a quickly growing piece of this market. As of 2020, the global text to speech market was valued at \$2 billion USD and is expected to grow to \$5 billion USD by 2026.<sup>10</sup>

With the popularization of smart speakers occurring within the general public and the widespread adoption of virtual assistants within businesses, synthetic voice is clearly heading into mainstream utility. It will be widely used, consistently iterated, and employed for a plethora of different purposes.

### The Benefits of Synthetic Voice

Synthetic voice can be used for extremely positive reasons. Two examples often cited include the ability of actor Val Kilmer to “speak” after throat cancer<sup>11</sup> and the firm Sonantic’s work archiving his voice. Below is a list of some of these positive use cases with inclusion and access for those with disability, adding additional layers to authentication of sensitive documents, and improving customer service systems such as IVR ranking as most significant.

Positive use cases are listed as follows:

- Inclusion and access for those with disabilities (i.e., the blind or the functionally illiterate) such as archiving voices lost such as Val Kilmer<sup>12</sup>
- Transposing existing text based digital content to speech (i.e., news and corporate publications and communication). (Here, an example of Trinity Audio’s text to audio player at work.)<sup>13</sup>
- Aiding in educational learning and content
- Advanced manufacturing manuals and FAQ
- Chatbots/voicebots
- Multiple voice options on one platform
- Audio books
- Gaming (An example found at the reference.<sup>14</sup>)
- Voice authentications for financial or sensitive materials. US Bank implemented in 2020 voice assistance as a part of their mobile banking options<sup>15</sup>
- Licensing and monetization for voice actors.
- Improving IVR and customer service systems (Statista<sup>16</sup> reported that organizations that adopted voice assistance decreased their customer service costs as well as the number of calls to their call centers by 20%)
- Preserving/archiving indigenous and or ethnic voices
- Podcasting with a more engaging synthetic voice. For example, Rumble Studio’s Audio Leads podcast demonstrates their asynchronous podcast recording product<sup>17</sup>
- Streamlining the cultivation and creation of content for companies and brands.

## Potential Threats and Harms of Synthetic Voice

There are many widely-known examples of synthetic voice being used for nefarious ends:

- Data breaches and identity theft: an example, the theft of \$35 million from a Dubai bank using AI Voice cloning in 2021.<sup>18</sup>
- The abuse and theft of financial or sensitive materials through false voice authentications such as the theft of \$243,000 by impersonating the CEO of an energy firm in 2019.<sup>19</sup>
- Spreading misleading information through false content; a society-wide worry.
- Falsified attribution and the illegal impersonation of voice actors and celebrities. This is theft of one’s intellectual property i.e., their voice. An example of this was the improper use of Bev Standing’s voice on TikTok.<sup>20</sup>
- Social bullying, not dissimilar from a Facebook comment rant, can affect any of us in a very hostile way. And hitting a record button is just as easy and fast as hitting a post button before thinking.
- Blurring the line between human talent and digital art. For example, NFT’s that are a recreation of already existing art. Some artists have argued that their artwork has been tokenized without their permission.<sup>21</sup>

## The Stakeholders in Synthetic Voice and How Are They Impacted?

From voice actors to advertisers to gamers, there is a wide range of individuals who may be impacted by the usage of synthetic voice. Here is a working list of potential stakeholders

- Voice actors/agencies/managers/unions (i.e., SAG-AFTRA)
  - Harm: theft of IP.
  - Benefit: new revenue streams.
- Synthetic voice creators and production companies
  - Harm: Spreading misinformation.
  - Benefit: New innovation and creation possibilities for a fast-growing voice tech world.
- Advertisers
  - Harm: Alienating customers with inauthentic advertising and marketing.
  - Benefit: New outlets created for reach.
- Gamers
  - Harm: Data breaches.
  - Benefit: Streamlining content and creating new creative play options.
- The general public (their belief/faith in the content that they consume)
  - Harm: Distrust of information and data breaches.
  - Benefit: Alternative interaction options in the day to day.
- Entertainment lawyers
  - Harm: Lack of education on synthetic voice can cause their clients to lose in many ways.
  - Benefit: With rules and standards in place, their clients could have more protection against misuse in the future.
- Musicians/composers/producers
  - Harm: Blurring lines of human talent.
  - Benefit: New revenue streams.
- Podcasters
  - Harm: Losing authenticity and connection with the audience.
  - Benefit: Streamlining content creation.
- Executives/newscasters/spokespeople
  - Harm: Spread of misinformation.
  - Benefit: Streamlining content and connecting with a new audience that does not read or reads far less than listens.
- Voiceover marketplaces (i.e., Voice123.com, CastVoices, Lotas Productions, etc.)
  - Harm: theft of IP.
  - Benefit: New revenue streams.

- Children and parents relative to their toys and manipulatives
  - Harm: Inappropriate representation of voice or inaccurate language/information.
  - Benefit: More ways of learning and playing.

## Next Steps: Development of Guidelines and Standards

How do we proceed? Below is a list of five action steps (in order of priority) that should help to begin the process. These steps will help to protect the use of synthetic voice and best position it to be of benefit to the population rather than a harm.

1. Education and awareness throughout the voice and synthetic voice industry and beyond. This would be executed through industry events, press events, and general social and social media awareness.
2. Creating a means of detecting fakes; a watermarking service or some kind of authenticator that can be turned on and off. This technology will give consumers the ability to detect fraudulent usage.
3. Governance and guidelines. Establishing rules and a body for governance (MPAA) to work toward something like a VAST template in advertising. This will help to assure that the future developers and synthetic voice users will not have an easy time abusing the technology.
4. Establishment of a registration ledger for those natural human voices used to better track fraudulent vs. legitimate usages of voice actors IP.
5. A distinction between professional and casual voice actors within the governance and corresponding guidelines. (With reference to step 3.)

## The Purpose of the Synthetic Voice Study Group

To clarify the legal and ethical position around the issue of Synthetic Voice production and usages as well as to create and implement actionable steps to detect and protect against fraudulent use. Synthetic Voice Messaging Paper

## The Role of the Open Voice Network

The Open Voice Network promotes the values of ownership, explicit transparency, and consent. It also protects the rights of intellectual property and content creation and ownership. The OVON seeks to make voice technology worthy of user trust through the communal development and adoption of industry standards and usage guidelines, industry education and advocacy initiatives, and the development and documentation of voice-centric value propositions.

## Glossary

**Synthetic Voice-** A machine generated replica of a natural human voice.

**Speech Synthesis-** The process of generating spoken language by machine based on written input. **Voice**

**Verification-** Utilizing voice as a multiple step verification process (i.e., authenticating identity through written and verbal authentication process).

**Speech Cloning-** The creation of an artificial simulation of a person's voice.

**Audio Deepfake-** When the voice has been digitally altered to sound like someone or something else, typically used maliciously or to spread false information.

## End Notes

---

1. <https://www.edisonresearch.com/the-infinite-dial-2021-2/>
2. <https://beyondwords.io/blog/speechkit-ai-audio-article-report-2021/>
3. [https://www.nationalpublicmedia.com/uploads/2021/11/The\\_Spoken\\_Word\\_Audio\\_Report\\_2021\\_For\\_Download.pdf](https://www.nationalpublicmedia.com/uploads/2021/11/The_Spoken_Word_Audio_Report_2021_For_Download.pdf)
4. <https://www.edisonresearch.com/the-infinite-dial-2021-2>
5. <https://www.statista.com/statistics/1133875>
6. <https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/> /global-voice-recognition-market-size/
7. <https://www.globenewswire.com/news-release/2021/04/13/2209306/0/en/Global-Voice-Cloning-Market-to-Reach-3-8-Billion-by-2027.html>
8. <https://www.meticulousresearch.com/pressrelease/280/speech-and-voice-recognition-market-2025>
9. <https://www.researchandmarkets.com/reports/5240193/global-text-to-speech-market-with-covid-19-impact>
10. <https://www.statista.com/topics/5572/virtual-assistants/>
11. <https://www.washingtonpost.com/technology/2021/08/18/val-kilmer-ai-voice-cloning/>
12. Ibid
13. [https://demo.trinityaudio.ai/tunein\\_openmode/](https://demo.trinityaudio.ai/tunein_openmode/)
14. <https://www.wanderword.net>
15. <https://voicebot.ai/2020/07/28/u-s-bank-launches-mobile-app-voice-assistant-built-to-mimic-human-bank-tellers/#:~:>
16. <https://www.statista.com/statistics/1073798/worldwide-high-benefit-voice-assistant-business>
17. <https://rumble.studio/podcast-show/audio-leads>
18. <https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/>?
19. <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>
20. <https://www.theverge.com/2021/9/29/22701167/bev-standing-tiktok-lawsuit-settles-text-to-speech-voice>
21. <https://www.jdsupra.com/legalnews/nfts-legal-risks-from-minting-art-and-4997056/>

## About the Open Voice Network

The Open Voice Network (OVON) is a non-profit industry association dedicated to the development of standards for voice assistance transparency, consent, limited collection, and control of voice data that will make voice technology worthy of user trust. The Open Voice Network operates as an open-source community of the Linux Foundation. It is independently governed and funded.